GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 07:55:03 ; Search time 1615 Seconds
(without alignments)
436.427 Million cell updates/sec

Title: US-09-310-844C-24
Perfect score: 29
Sequence: 1 uaugauucuuuuuguaagcccuaggggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 243536

Minimum DB seq length: 0
Maximum DB seq length: 70

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	16.2	55.9	57	9	AI561770	AI561770 vv65b08.x
c	2	15.4	53.1	66	29	TA123H02P	AL463084 T. brucei
c	3	15.2	52.4	41	14	CB210991	CB210991 OML01271
	4	15.2	52.4	44	9	AL892747	AL892747 AL892747
	5	15.2	52.4	63	29	BX533881	BX533881 Arabidops
c	6	14.8	51.0	25	28	AZ993079	AZ993079 2M0277P20
	7	14.8	51.0	60	10	BE871815	BE871815 601447803
c	8	14.8	51.0	64	29	BX003595	BX003595 Arabidops
	9	14.8	51.0	68	9	AL960604	AL960604 AL960604
	10	14.8	51.0	69	12	BM128463	BM128463 if15c05.x
	11	14.6	50.3	49	29	BX287070	BX287070 Arabidops
	12	14.6	50.3	52	10	BG236504	BG236504 nai44c06.
c	13	14.6	50.3	62	28	BH911891	BH911891 SALK_0727
c	14	14.6	50.3	64	9	AI139668	AI139668 qc28h07.x
	15	14.6	50.3	64	28	AZ808107	AZ808107 2M0071O24
	16	14.6	50.3	66	28	AZ440181	AZ440181 1M0231E10
c	17	14.4	49.7	31	28	BH910631	BH910631 SALK_0607
c	18	14.4	49.7	43	28	AZ597048	AZ597048 1M0410K10
	19	14.4	49.7	44	29	AL771575	AL771575 Arabidops
	20	14.4	49.7	54	12	BI865449	BI865449 ft22h02.x
	21	14.4	49.7	60	9	AL595218	AL595218 AL595218
c	22	14.4	49.7	63	10	BG362434	BG362434 gb72b09.y
c	23	14.4	49.7	67	9	AI584052	AI584052 ts13b02.x
c	24	14.4	49.7	67	10	BE027305	BE027305 EtESTea89
	25	14.4	49.7	67	29	BZ289657	BZ289657 SALK_0230
c	26	14.4	49.7	69	10	BE647308	BE647308 UI-M-BH1-
	27	14.2	49.0	52	9	AW686481	AW686481 NF038E05N
c	28	14.2	49.0	59	28	B00509	B00509 cSRL-115b2-
	29	14.2	49.0	61	9	AI318033	AI318033 ta75g02.x
	30	14.2	49.0	67	29	BX004510	BX004510 Arabidops
	31	14.2	49.0	70	13	BU063954	BU063954 Fgr_3_M18
c	32	14	48.3	52	10	BF637245	BF637245 NF002F08L
c	33	14	48.3	65	12	BI094834	BI094834 EST-CD34N
c	34	14	48.3	67	9	AA936041	AA936041 nz53f10.s
	35	14	48.3	67	28	BH855810	BH855810 SALK_0845
	36	13.8	47.6	37	29	AL951243	AL951243 Arabidops
	37	13.8	47.6	39	28	BH909815	BH909815 SALK_0561
	38	13.8	47.6	40	28	BH857340	BH857340 SALK_0764
	39	13.8	47.6	40	28	BH857342	BH857342 SALK_0764
	40	13.8	47.6	43	28	AZ484548	AZ484548 1M0311N02
	41	13.8	47.6	55	9	AT005996	AT005996 AT005996
c	42	13.8	47.6	55	9	AA276988	AA276988 va25g09.r
	43	13.8	47.6	64	29	AL770628	AL770628 Arabidops
c	44	13.8	47.6	65	29	AL763793	AL763793 Arabidops
c	45	13.8	47.6	66	29	AL767936	AL767936 Arabidops

ALIGNMENTS

RESULT 1
AI561770
LOCUS AI561770 57 bp mRNA linear EST 25-MAR-1999
DEFINITION vv65b08.x1 Stratagene mouse skin (#937313) Mus musculus cDNA clone IMAGE:1227255 3', mRNA sequence.
ACCESSION AI561770
VERSION AI561770.1 GI:4513115
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 57)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the

```
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:652847
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            High quality sequence stop: 51.
FEATURES             Location/Qualifiers
     source          1. .57
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1227255"
                     /sex="females"
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
                     ; Site_2: XhoI; Cloned unidirectionally.  Primer: Oligo
                     dT. Whole skin from 11 week old C57BL/6 female mice.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       17 a      9 c     12 g     19 t
ORIGIN

  Query Match             55.9%;  Score 16.2;  DB 9;  Length 57;
  Best Local Similarity   37.9%;  Pred. No. 6.7e+04;
  Matches   11;  Conservative   10;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
              : :|| :| :::: :|| || : |||| :
Db         25 TTTGAATCCTTTTCTAATCCATGGGGGGT 53


RESULT 2
TA123H02P/c
LOCUS       TA123H02P                 66 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 123h02, forward sequence,
            genomic survey sequence.
ACCESSION   AL463084
VERSION     AL463084.1  GI:11833690
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 66)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh1@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1. .66
                     /organism="Trypanosoma brucei"
                     /mol_type="genomic DNA"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="123h02"
BASE COUNT       29 a      8 c      9 g     20 t
ORIGIN

  Query Match             53.1%;  Score 15.4;  DB 29;  Length 66;
  Best Local Similarity   36.0%;  Pred. No. 1.3e+05;
  Matches    9;  Conservative   10;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGG 25
              :|:||:: :::::   | |||:| |
Db         30 TATGATTTTTTTTCAGAACCCTAAG 6


RESULT 3
CB210991/c
LOCUS       CB210991                  41 bp    mRNA    linear   EST 05-FEB-2003
DEFINITION  OML01271 Oryza minuta HybriZAP-2.1 XR library Oryza minuta cDNA 5',
            mRNA sequence.
ACCESSION   CB210991
VERSION     CB210991.1  GI:28257082
KEYWORDS    EST.
SOURCE      Oryza minuta
  ORGANISM  Oryza minuta
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 41)
  AUTHORS   Shin,J.S.
  TITLE     Oryza minuta HybriZAP-2.1 XR library
  JOURNAL   Unpublished
COMMENT     Contact: Jeong Sheop Shin
            Plant Molecular Genetics
            Graduate School of Biotechnology, University of Korea
            136-701 Anam-dong 5/1 Seoul, Korea
            Tel: 00 82 2 3290 3430
            Fax: 00 82 2 927 9028
            Email: jsshin@kuccnx.korea.ac.kr.
FEATURES             Location/Qualifiers
     source          1. .41
                     /organism="Oryza minuta"
                     /mol_type="mRNA"
                     /db_xref="taxon:63629"
                     /dev_stage="4-weeks after germination"
                     /clone_lib="Oryza minuta HybriZAP-2.1 XR library"
                     /note="Organ: immature leaf"
BASE COUNT       18 a     10 c     11 g      2 t
ORIGIN

  Query Match             52.4%;  Score 15.2;  DB 14;  Length 41;
  Best Local Similarity   39.3%;  Pred. No. 1.6e+05;
  Matches   11;  Conservative    9;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGC 28
              :|:| ::|:: :  : |||||: | | |
Db         28 TATGCTTCTTGTGCTTAGCCCTCGTGCC 1


RESULT 4
AL892747
LOCUS       AL892747                  44 bp    mRNA    linear   EST 16-SEP-2002
DEFINITION  AL892747 XGC-egg Silurana tropicalis cDNA clone TEgg041m06 5', mRNA
            sequence.
ACCESSION   AL892747
VERSION     AL892747.1  GI:22943298
KEYWORDS    EST.
SOURCE      Silurana tropicalis (western clawed frog)
  ORGANISM  Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE   1  (bases 1 to 44)
  AUTHORS   Taylor,R., Ashurst,J.L., Croning,M.D.R., Zorn,A.M. and Rogers,J.
```

```
  TITLE     Sanger Xenopus tropicalis EST project 2002
  JOURNAL   Unpublished
COMMENT     Contact: Taylor R
            Sanger Centre
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: trop@sanger.ac.uk
            Sanger Xenopus tropicalis EST project 2001
            TROPICALIS_SEQUENCE_ID: TEgg041m06.p1kSP6
            Sequencing primer: SP6
            This sequence is from a Xenopus Gene Collection (XGC) library
            constructed by Aaron M. Zorn.
FEATURES             Location/Qualifiers
     source          1. .44
                     /organism="Silurana tropicalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8364"
                     /clone="TEgg041m06"
                     /dev_stage="egg"
                     /lab_host="Escherichia coli XL1-blue"
                     /clone_lib="XGC-egg"
                     /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
                     was oligo dT primed from 5ug of poly A+ RNA from egg.
                     EcoRI-NotI cut cDNA was then ligated into pCS107 with
                     EcoRI at the 5' end and NotI at the 3' end"
BASE COUNT       10 a      7 c     10 g     16 t      1 others
ORIGIN

  Query Match             52.4%;  Score 15.2;  DB 9;  Length 44;
  Best Local Similarity   39.3%;  Pred. No. 1.6e+05;
  Matches   11;  Conservative    9;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGC 28
              :|: | :|::: :|:   || :|||| |
Db          3 TATAAGTCTTTATGTTCTCCATAGGGAC 30


RESULT 5
BX533881
LOCUS       BX533881                  63 bp    DNA     linear   GSS 03-JUN-2003
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-504G04-019705,
            genomic survey sequence.
ACCESSION   BX533881
VERSION     BX533881.1  GI:31411011
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.
            and Weisshaar,B.
  TITLE     A pipeline for automated high-throughput generation of FSTs
            (flanking sequence tags) from Arabidopsis thaliana T-DNA
            transformed lines
  JOURNAL   Unpublished
REFERENCE   2
  AUTHORS   Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
  TITLE     A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
            for flanking sequence tag based reverse genetics
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 63)
  AUTHORS   Strizhov,N., Li,Y., Rosso,M. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2003) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence is recovered from the left border of the T-DNA. It
            indicates an insertion close to or within gene At5g40230. The
            sequences are generated at the MPI for Plant Breeding Research in
            the context of the GABI-Kat project. GABI-Kat is part of the German
            Plant Genomics program designated 'GABI'. Information on line
            availability can be found at:
            http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .63
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="GK-504G04-019705"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
                     vector pAC161. The lines contain one or more T-DNA
                     insertions. The DNA fragment(s) resulting from the PCR
                     were directly sequenced to determine the genomic sequence
                     flanking the insertion. Sequences displaying significant
                     similarity to the A. thalaina nuclear genome sequence were
                     processed for submission. T-DNA derived sequences were
                     removed"
BASE COUNT       18 a     15 c     12 g     18 t
ORIGIN

  Query Match             52.4%;  Score 15.2;  DB 29;  Length 63;
  Best Local Similarity   39.3%;  Pred. No. 1.6e+05;
  Matches   11;  Conservative    9;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          2 AUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
              |: | : :::::  || |||:| ||| :
Db         21 ATAAATATTTTTTCAAACCCTATGGGAT 48


RESULT 6
AZ993079/c
LOCUS       AZ993079                  25 bp    DNA     linear   GSS 27-APR-2001
DEFINITION  2M0277P20R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0277P20 R, genomic survey sequence.
ACCESSION   AZ993079
VERSION     AZ993079.1  GI:13864306
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0277  row: P  column: 20
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 25.
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0277P20"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
```

musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```
BASE COUNT        4 a      8 c      6 g      7 t
ORIGIN

  Query Match            51.0%;  Score 14.8;  DB 28;  Length 25;
  Best Local Similarity   72.2%;  Pred. No. 2.3e+05;
  Matches   13;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         11 UUUGUAAGCCCUAGGGGC 28
              :::| |||||| ||||||
Db         21 TTTGCAAGCCCAAGGGGC 4


RESULT 7
BE871815
LOCUS       BE871815                 60 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601447803F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3851880 5',
            mRNA sequence.
ACCESSION   BE871815
VERSION     BE871815.1  GI:10320591
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 60)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9573  row: e  column: 01
            High quality sequence stop: 60.
FEATURES             Location/Qualifiers
     source          1. .60
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3851880"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_65"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.8 kb.   Library constructed by Life
                     Technologies.  "
BASE COUNT       16 a     11 c     10 g     23 t
ORIGIN

  Query Match            51.0%;  Score 14.8;  DB 10;  Length 60;
  Best Local Similarity   38.9%;  Pred. No. 2.2e+05;
  Matches    7;  Conservative    9;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 AUGAUUCUUUUUGUAAGC 19
              |:||:: ::::: :||||
Db         10 ATGATTATTTTTCTAAGC 27


RESULT 8
BX003595/c
LOCUS       BX003595                 64 bp    DNA     linear   GSS 04-DEC-2002
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-373C10-017165,
            genomic survey sequence.
ACCESSION   BX003595
VERSION     BX003595.1  GI:26188555
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.
            and Weisshaar,B.
  TITLE     A pipeline for automated high-throughput generation of FSTs
            (flanking sequence tags) from Arabidopsis thaliana T-DNA
            transformed lines
  JOURNAL   Unpublished
REFERENCE   2
  AUTHORS   Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
  TITLE     A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
            for flanking sequence tag based reverse genetics
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 64)
  AUTHORS   Strizhov,N., Li,Y., Rosso,M. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2002) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence is recovered from the left border of the T-DNA. It
            indicates an insertion close to or within gene At1g33610. The
            sequences are generated at the MPI for Plant Breeding Research in
            the context of the GABI-Kat project. GABI-Kat is part of the German
            Plant Genomics program designated 'GABI'. Information on line
            availability can be found at:
            http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .64
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="GK-373C10-017165"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
                     vector pAC161. The lines contain one or more T-DNA
                     insertions. The DNA fragment(s) resulting from the PCR
                     were directly sequenced to determine the genomic sequence
                     flanking the insertion. Sequences displaying significant
                     similarity to the A. thalaina nuclear genome sequence were
                     processed for submission. T-DNA derived sequences were
                     removed"
BASE COUNT       25 a     10 c     10 g     19 t
ORIGIN

  Query Match            51.0%;  Score 14.8;  DB 29;  Length 64;
  Best Local Similarity   34.6%;  Pred. No. 2.2e+05;
  Matches    9;  Conservative   10;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 AUGAUUCUUUUUGUAAGCCCUAGGGG 27
              |:  :: ::::: :| ||| : ||||
Db         60 ATCTTTTTTTTTATATGCCTTTGGGG 35
```

```
RESULT 9
AL960604
LOCUS       AL960604                  68 bp    mRNA    linear   EST 27-NOV-2002
DEFINITION  AL960604 XGC-gastrula Silurana tropicalis cDNA clone TGas120b03 5',
            mRNA sequence.
ACCESSION   AL960604
VERSION     AL960604.1  GI:25784199
KEYWORDS    EST.
SOURCE      Silurana tropicalis (western clawed frog)
  ORGANISM  Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE   1  (bases 1 to 68)
  AUTHORS   Taylor,R., Ashurst,J.L., Croning,M.D.R., Zorn,A.M. and Rogers,J.
  TITLE     Sanger Xenopus tropicalis EST project 2002
  JOURNAL   Unpublished
COMMENT     Contact: Taylor R
            Sanger Centre
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: trop@sanger.ac.uk
            Sanger Xenopus tropicalis EST project 2001
            TROPICALIS_SEQUENCE_ID: TGas120b03.p1kSP6
            Sequencing primer: SP6
            This sequence is from a Xenopus Gene Collection (XGC) library
            constructed by Aaron M. Zorn.
FEATURES             Location/Qualifiers
     source          1. .68
                     /organism="Silurana tropicalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8364"
                     /clone="TGas120b03"
                     /dev_stage="gastrula (stages 10.5-13 mixed)"
                     /lab_host="Escherichia coli XL1-blue"
                     /clone_lib="XGC-gastrula"
                     /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
                     was oligo dT primed from 5ug of poly A+ RNA from stages
                     10-13 gastrulae.  EcoRI-NotI cut cDNA was then ligated
                     into pCS107 with EcoRI at the 5' end and NotI at the 3'
                     end."
BASE COUNT        8 a     25 c     16 g     19 t
ORIGIN

  Query Match             51.0%;  Score 14.8;  DB 9;  Length 68;
  Best Local Similarity   34.6%;  Pred. No. 2.2e+05;
  Matches    9;  Conservative   10;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCUAGGGGCU 29
              | :: ::::: :|| |||:  || |:
Db         11 GGTTTTTTTTTTAACCCCTGTGGCCT 36


RESULT 10
BM128463
LOCUS       BM128463                  69 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  if15c05.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA clone IMAGE:5676297 3', mRNA sequence.
ACCESSION   BM128463
VERSION     BM128463.1  GI:17123015
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 69)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu) This sequence now available from the IMAGE
            consortium, for clone orders contact: info@image.llnl.gov.
FEATURES             Location/Qualifiers
     source          1. .69
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5676297"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit  (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by  column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified  once on solid support and plasmid DNA from
                     library was prepared. The  library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares  1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library  inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized)  plasmids were isolated
                     by hydroxyapatite chromatography and used to make  this
                     library."
BASE COUNT        8 a     10 c     19 g     32 t
ORIGIN

  Query Match             51.0%;  Score 14.8;  DB 12;  Length 69;
  Best Local Similarity   30.8%;  Pred. No. 2.2e+05;
  Matches    8;  Conservative   11;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGG 26
              : :  :: ::::: :  ||||:||||
Db          8 TTTTTTTTTTTTTCTGGGCCCTAGGG 33


RESULT 11
BX287070
LOCUS       BX287070                  49 bp    DNA     linear   GSS 07-MAR-2003
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-396F11-018295,
            genomic survey sequence.
ACCESSION   BX287070
VERSION     BX287070.1  GI:28886066
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.
            and Weisshaar,B.
  TITLE     A pipeline for automated high-throughput generation of FSTs
            (flanking sequence tags) from Arabidopsis thaliana T-DNA
            transformed lines
  JOURNAL   Unpublished
REFERENCE   2
```

```
  AUTHORS   Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
  TITLE     A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
            for flanking sequence tag based reverse genetics
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 49)
  AUTHORS   Strizhov,N., Li,Y., Rosso,M. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2003) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence is recovered from the left border of the T-DNA. It
            indicates an insertion within the locus defined by clone MZN24. The
            sequences are generated at the MPI for Plant Breeding Research in
            the context of the GABI-Kat project. GABI-Kat is part of the German
            Plant Genomics program designated 'GABI'. Information on line
            availability can be found at:
            http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="GK-396F11-018295"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
                     vector pAC161. The lines contain one or more T-DNA
                     insertions. The DNA fragment(s) resulting from the PCR
                     were directly sequenced to determine the genomic sequence
                     flanking the insertion. Sequences displaying significant
                     similarity to the A. thalaina nuclear genome sequence were
                     processed for submission. T-DNA derived sequences were
                     removed"
BASE COUNT       14 a      7 c     13 g     15 t
ORIGIN

  Query Match             50.3%;  Score 14.6;  DB 29;  Length 49;
  Best Local Similarity   41.4%;  Pred. No. 2.7e+05;
  Matches   12;  Conservative    8;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
              :| | ::   : :|:||| ||:| ||| :
Db          8 TAAGCTTAAATGTGTAAGGCCTATGGGAT 36


RESULT 12
BG236504
LOCUS       BG236504                 52 bp    mRNA    linear   EST 12-FEB-2001
DEFINITION  nai44c06.x1 NCI_CGAP_HN20 Homo sapiens cDNA clone IMAGE:4262795 3',
            mRNA sequence.
ACCESSION   BG236504
VERSION     BG236504.1  GI:12750351
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 52)
  AUTHORS   NCI/NIDR-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Dental Research,
            Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .52
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4262795"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_HN20"
                     /note="Organ: normal head/neck tissue; Vector: pAMP1; mRNA
                     made from head/neck tissue, cDNA made by oligo-dT
                     priming. Directionally cloned into UDG sites.
                     Size-selected on  agarose gel, average insert size 300 bp.
                     Primary library.  cDNA Library Preparation: David B.
                     Krizman, Ph.D."
BASE COUNT       11 a      9 c     10 g     22 t
ORIGIN

  Query Match             50.3%;  Score 14.6;  DB 10;  Length 52;
  Best Local Similarity   38.1%;  Pred. No. 2.7e+05;
  Matches    8;  Conservative    9;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          6 UUCUUUUUGUAAGCCCUAGGG 26
              :: ::::: :|||  |:||||
Db          1 TTTTTTTTTTAAGTTCTAGGG 21


RESULT 13
BH911891/c
LOCUS       BH911891                 62 bp    DNA     linear   GSS 04-SEP-2002
DEFINITION  SALK_072729.50.65.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_072729.50.65.x, genomic
            survey sequence.
ACCESSION   BH911891
VERSION     BH911891.1  GI:22724824
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 62)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA. This sequence lies within 300 bases of the 5' end of
            At4g13720.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .62
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_072729.50.65.x"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
BASE COUNT       28 a      7 c     12 g     15 t
ORIGIN
```

```
  Query Match             50.3%;  Score 14.6;  DB 28;  Length 62;
  Best Local Similarity   31.0%;  Pred. No. 2.7e+05;
  Matches    9;  Conservative   11;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
              :|: |:: ::::  :| || |:||   |:
Db         46 TATAATTTTTTTACTAGGCGCTAGACCCT 18


RESULT 14
AI139668/c
LOCUS       AI139668                 64 bp    mRNA    linear   EST 27-OCT-1998
DEFINITION  qc28h07.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:1710973 3', mRNA sequence.
ACCESSION   AI139668
VERSION     AI139668.1  GI:3645640
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 64)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1168   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .64
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1710973"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Soares_pregnant_uterus_NbHPU"
                     /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization.  Library
                     constructed by M. Fatima Bonaldo."
BASE COUNT       24 a     13 c      5 g     22 t
ORIGIN

  Query Match             50.3%;  Score 14.6;  DB 9;  Length 64;
  Best Local Similarity   34.5%;  Pred. No. 2.7e+05;
  Matches   10;  Conservative   10;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
              : : |:: : ::: :||  || ||||| :
Db         48 TTTAATTTTATTTTTAATTCCCAGGGGTT 20


RESULT 15
AZ808107
LOCUS       AZ808107                 64 bp    DNA     linear   GSS 20-FEB-2001
DEFINITION  2M0071O24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0071O24 F, genomic survey sequence.
ACCESSION   AZ808107
VERSION     AZ808107.1  GI:12973121
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 64)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ;M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0071  row: O  column: 24
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 64.
FEATURES             Location/Qualifiers
     source          1. .64
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0071O24"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        9 a     10 c      7 g     38 t
ORIGIN

  Query Match             50.3%;  Score 14.6;  DB 28;  Length 64;
  Best Local Similarity   27.6%;  Pred. No. 2.7e+05;
  Matches    8;  Conservative   12;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
              : :  :: ::::: : |||  : |||||:
Db         16 TTTTTTTTTTTTTTTAGCAATTGGGGCT 44


Search completed: January 30, 2004, 10:12:36
Job time : 1623 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 06:57:47 ; Search time 575.333 Seconds
(without alignments)
2062.073 Million cell updates/sec

Title: US-09-310-844C-25
Perfect score: 29
Sequence: 1 aaagauucuuuuuguaagccccaagggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 1427288

Minimum DB seq length: 0
Maximum DB seq length: 70

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18.4	63.4	43	8	ATH553793	AJ553793 Arabidops
2	c	16.4	56.6	31	6	AX425989	AX425989 Sequence
3	c	16.2	55.9	48	6	AX018731	AX018731 Sequence
4	c	15.2	52.4	33	6	AR020509	AR020509 Sequence
5		15.2	52.4	47	6	AR289362	AR289362 Sequence
6		15	51.7	23	6	E09974	E09974 Primer for
7		15	51.7	23	6	E10118	E10118 PCR primer
8	c	15	51.7	33	6	BD094291	BD094291 A prepara
9	c	15	51.7	33	6	BD179370	BD179370 A novel p
10	c	15	51.7	41	6	AX514720	AX514720 Sequence
11	c	15	51.7	41	6	AX520728	AX520728 Sequence
12	c	14.8	51.0	29	6	AR019319	AR019319 Sequence
13	c	14.8	51.0	29	6	AR061847	AR061847 Sequence
14	c	14.8	51.0	29	6	AR147578	AR147578 Sequence
15	c	14.8	51.0	29	6	AR252838	AR252838 Sequence
16	c	14.8	51.0	29	6	I34733	I34733 Sequence 25
17	c	14.8	51.0	29	6	I67987	I67987 Sequence 25
18	c	14.8	51.0	32	6	AR061867	AR061867 Sequence
19	c	14.8	51.0	32	6	AR252858	AR252858 Sequence
20		14.6	50.3	25	6	AR206010	AR206010 Sequence
21		14.6	50.3	25	6	AX043294	AX043294 Sequence
22	c	14.6	50.3	53	6	AR061021	AR061021 Sequence
23		14.4	49.7	25	6	AX043671	AX043671 Sequence
24	c	14.4	49.7	31	6	AX425978	AX425978 Sequence
25		14.4	49.7	51	6	AX117185	AX117185 Sequence
26		14.4	49.7	59	11	AL772710	AL772710 Arabidops
27		14.4	49.7	59	11	AL773099	AL773099 Arabidops
28	c	14.2	49.0	31	6	AX582577	AX582577 Sequence
29	c	14.2	49.0	44	6	AX601758	AX601758 Sequence
30		14.2	49.0	47	6	AR288361	AR288361 Sequence
31		14.2	49.0	65	6	AX483200	AX483200 Sequence
32		14.2	49.0	69	6	AR052906	AR052906 Sequence
33		14.2	49.0	69	6	AR054269	AR054269 Sequence
34		14.2	49.0	69	6	AR054471	AR054471 Sequence
35		14.2	49.0	69	8	ATH527686	AJ527686 Arabidops
36		14	48.3	25	6	AX043055	AX043055 Sequence
37	c	14	48.3	29	11	AL806194	AL806194 Arabidops
38	c	14	48.3	29	11	AL824524	AL824524 Arabidops
39	c	14	48.3	33	11	AL806370	AL806370 Arabidops
40	c	14	48.3	34	11	AL824530	AL824530 Arabidops
41	c	14	48.3	35	11	AL806140	AL806140 Arabidops
42	c	14	48.3	35	11	AL806354	AL806354 Arabidops
43	c	14	48.3	40	11	AL824626	AL824626 Arabidops
44	c	14	48.3	40	11	AL824634	AL824634 Arabidops
45		14	48.3	41	6	AR253879	AR253879 Sequence

ALIGNMENTS

RESULT 1
ATH553793
LOCUS ATH553793 43 bp DNA linear PLN 29-MAR-2003
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, left border, clone 383B09.
ACCESSION AJ553793
VERSION AJ553793.1 GI:29370260
KEYWORDS left border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,

```
            Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G.,
            Lepiniec,L., Caboche,M. and Lecharny,A.
  TITLE     T-DNA integration into the Arabidopsis genome depends on sequences
            of pre-insertion sites
  JOURNAL   EMBO Rep. 3 (12), 1152-1157 (2002)
  MEDLINE   22363535
   PUBMED   12446565
REFERENCE   2  (bases 1 to 43)
  AUTHORS   Balzergue,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue
            Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .43
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Wassillewskija"
                     /db_xref="taxon:3702"
                     /clone="383B09"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
     misc_feature    1. .43
                     /note="T-DNA flanking sequence
                     left border"
BASE COUNT       15 a      7 c      4 g     17 t
ORIGIN

  Query Match           63.4%;  Score 18.4;  DB 8;  Length 43;
  Best Local Similarity   50.0%;  Pred. No. 1e+03;
  Matches   14;  Conservative    8;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AAAGAUUCUUUUUGUAAGCCCCAAGGGC 28
              ||| |::|:::::|:| ||  |||  ||
Db          2 AAAAATTCTTTTTGTAGGCATCAAATGC 29


RESULT 2
AX425989/c
LOCUS       AX425989                  31 bp    DNA     linear   PAT 18-JUN-2002
DEFINITION  Sequence 4325 from Patent WO0188124.
ACCESSION   AX425989
VERSION     AX425989.1  GI:21529375
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Jarvis,T., von Carlowitz,I., Mcswiggen,J.A., Mclaughlin,F.G. and
            Randi,A.M.
  TITLE     Method and reagent for the inhibition of erg
  JOURNAL   Patent: WO 0188124-A 4325 22-NOV-2001;
            RIBOZYME PHARMACEUTICALS, INC. (US) ; GLAXO GROUP LIMITED (GB)
FEATURES             Location/Qualifiers
     source          1. .31
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Enzymatic Nucleic Acid"
BASE COUNT        6 a      10 c      8 g      7 t
ORIGIN

  Query Match           56.6%;  Score 16.4;  DB 6;  Length 31;
  Best Local Similarity   61.5%;  Pred. No. 9.4e+03;
  Matches   16;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 AGAUUCUUUUUGUAAGCCCCAAGGGC 28
              |||:  :: : |  ||||||||||||
Db         26 AGATCGTTGTAGCTAGCCCCAAGGGC 1


RESULT 3
AX018731/c
LOCUS       AX018731                  48 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 20 from Patent WO9944633.
ACCESSION   AX018731
VERSION     AX018731.1  GI:10042853
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Minke,J.M. and Audonnet,J.C.
  TITLE     Live recombined vaccines injected with adjuvant
  JOURNAL   Patent: WO 9944633-A 20 10-SEP-1999;
            MINKE JULES MAARTEN (FR); MERIAL SAS (FR); AUDONNET JEAN CHRISTOPHE
            FRANC (FR)
FEATURES             Location/Qualifiers
     source          1. .48
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="oligonucleotide"
BASE COUNT       17 a      9 c      9 g     13 t
ORIGIN

  Query Match           55.9%;  Score 16.2;  DB 6;  Length 48;
  Best Local Similarity   44.8%;  Pred. No. 1.2e+04;
  Matches   13;  Conservative    8;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 29
              |||  :  :::::|:|||| | ||||:
Db         37 AAATCTAATTTTTGTAAGCTTCCCGGGCT 9


RESULT 4
AR020509/c
LOCUS       AR020509                  33 bp    DNA     linear   PAT 05-DEC-1998
DEFINITION  Sequence 5 from patent US 5789171.
ACCESSION   AR020509
VERSION     AR020509.1  GI:3975124
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 33)
  AUTHORS   Smeltzer,M.S.
  TITLE     Use of cna, fnba, fnbb, and hlb, gene probes for the
            strain-specific identification of Staphylococcus aureus
  JOURNAL   Patent: US 5789171-A 5 04-AUG-1998;
FEATURES             Location/Qualifiers
     source          1. .33
                     /organism="unknown"
BASE COUNT       12 a      8 c      7 g      6 t
ORIGIN

  Query Match           52.4%;  Score 15.2;  DB 6;  Length 33;
  Best Local Similarity   42.9%;  Pred. No. 3.5e+04;
  Matches   12;  Conservative    8;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          2 AAGAUUCUUUUUGUAAGCCCCAAGGGCU 29
              | ||:: :::: |:||   ||  ||||:
Db         32 ATGATTGTTTTAGTAATTTCCCCGGGCT 5


RESULT 5
AR289362
```

```
LOCUS       AR289362                 47 bp    DNA     linear   PAT 12-JUN-2003
DEFINITION  Sequence 1097 from patent US 6537751.
ACCESSION   AR289362
VERSION     AR289362.1  GI:31676646
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 47)
  AUTHORS   Cohen,D., Chumakov,I. and Blumenfeld,M.
  TITLE     Biallelic markers for use in constructing a high density
            disequilibrium map of the human genome
  JOURNAL   Patent: US 6537751-A 1097 25-MAR-2003;
FEATURES             Location/Qualifiers
     source          1. .47
                     /organism="unknown"
BASE COUNT       13 a      9 c     10 g     14 t      1 others
ORIGIN

  Query Match            52.4%;  Score 15.2;  DB 6;  Length 47;
  Best Local Similarity  50.0%;  Pred. No. 3.5e+04;
  Matches   11;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          3 AGAUUCUUUUUGUAAGCCCCAA 24
              ||| :|::::|:|| | || |
Db         15 AGACTCTTTWTGTAACCTCCCA 36


RESULT 6
E09974
LOCUS       E09974                   23 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  Primer for amplifying human herpes virus.
ACCESSION   E09974
VERSION     E09974.1  GI:22026598
KEYWORDS    JP 1995250699-A/20.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 23)
  AUTHORS   Yamanishi,K., Mukai,T., Aono,T., Kondo,M. and Takarada,Y.
  TITLE     METHOD FOR DISCRIMINATORY DETECTION OF HUMAN HERPES VIRUS AND
            REAGENT THEREFOR
  JOURNAL   Patent: JP 1995250699-A 20 03-OCT-1995;
            TOYOBO CO LTD
COMMENT     OS   None
            OC   Artificial sequences.
            PN   JP 1995250699-A/20
            PD   03-OCT-1995
            PF   11-MAR-1994 JP 1994041101
            PI   YAMANISHI KOICHI, MUKAI TORU, AONO TOSHIYA, KONDO MOTOHIRO, PI
              TAKARADA YUTAKA
            PC   C12Q1/68,C12N15/09,C12Q1/70;
            CC   strandedness: Single;
            CC   topology: Linear;
            CC   hypothetical: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .23
            FT                   /organism='Artificial sequences' FT
            misc_feature    1. .23
            FT                   /note='Common sequences for human herpes virus
             FT                   6-type A,
            FT                   human herpes virus 6-type B,human herpes  FT
                            virus-7 and
            FT                   cytomegalovirus'.
FEATURES             Location/Qualifiers
     source          1. .23
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT       5 a      5 c      6 g      7 t
ORIGIN

  Query Match            51.7%;  Score 15;  DB 6;  Length 23;
  Best Local Similarity  52.2%;  Pred. No. 4.4e+04;
  Matches   12;  Conservative    6;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGG 26
              ||: |:::::|. |  |||||||
Db          1 GATCCTTTTTGAGATGCCCAAGG 23


RESULT 7
E10118
LOCUS       E10118                   23 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  PCR primer to detect human herpes virus-6A,6B and 7.
ACCESSION   E10118
VERSION     E10118.1  GI:22026746
KEYWORDS    JP 1995284391-A/6.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 23)
  AUTHORS   Yamanishi,K., Mukai,T., Aono,T., Kondo,M. and Takarada,Y.
  TITLE     OLIGONUCLEOTIDE FOR DETECTING HUMAN HERPES VIRUS AND ITS USE
  JOURNAL   Patent: JP 1995284391-A 6 31-OCT-1995;
            TOYOBO CO LTD
COMMENT     OS   None
            OC   Artificial sequences.
            PN   JP 1995284391-A/6
            PD   31-OCT-1995
            PF   19-APR-1994 JP 1994080488
            PI   YAMANISHI KOICHI, MUKAI TORU, AONO TOSHIYA, KONDO MOTOHIRO, PI
              TAKARADA YUTAKA
            PC   C12N15/09,C07H21/04,C12Q1/68,G01N33/567,G01N33/569; CC
            strandedness: Single;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .23
            FT                   /organism='Artificial sequences'.
FEATURES             Location/Qualifiers
     source          1. .23
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT       5 a      5 c      6 g      7 t
ORIGIN

  Query Match            51.7%;  Score 15;  DB 6;  Length 23;
  Best Local Similarity  52.2%;  Pred. No. 4.4e+04;
  Matches   12;  Conservative    6;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGG 26
              ||: |:::::|  |  |||||||
Db          1 GATCCTTTTTGAGATGCCCAAGG 23


RESULT 8
BD094291/c
LOCUS       BD094291                 33 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  A preparation method of adrenomedullin precursor.
ACCESSION   BD094291
VERSION     BD094291.1  GI:22639879
KEYWORDS    WO 0127310-A/41.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 33)
  AUTHORS   Takimoto,A., Mitsuda,Y., Nakayama,T. and Mitsushima,K.
  TITLE     A preparation method of adrenomedullin precursor
  JOURNAL   Patent: WO 0127310-A 41 19-APR-2001;
            SHIONOGI & CO LTD,AKIO TAKIMOTO,YUICHI MITSUDA,TOSHIMASA NAKAYAMA,
            KENJI MITSUSHIMA
COMMENT     OS   Artificial Sequence
            PN   WO 0127310-A/41
```

```
            PD   19-APR-2001
            PF   10-OCT-2000 WO 2000JP007023
            PR   15-OCT-1999 JP 99P   294147
            PI   AKIO TAKIMOTO,YUICHI MITSUDA,TOSHIMASA NAKAYAMA,KENJI  PI
           MITSUSHIMA
            PC   C12P21/02,C12N15/09,C12N15/12,C07K14/47,C07K19/00,C07K1/12, PC
             C07K1/30
            CC   Description of Artificial Sequence:Sense primer for HdeB FH
            Key             Location/Qualifiers.
FEATURES              Location/Qualifiers
     source           1. .33
                      /organism="synthetic construct"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:32630"
BASE COUNT       10 a      8 c      4 g     11 t
ORIGIN

  Query Match            51.7%;  Score 15;  DB 6;  Length 33;
  Best Local Similarity   47.8%;  Pred. No. 4.4e+04;
  Matches   11;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 AAAGAUUCUUUUUGUAAGCCCCA 23
              ||| |::| ::::|:|| |  ||
Db         24 AAATATTCATTTTGTAACCTGCA 2


RESULT 9
BD179370/c
LOCUS       BD179370                  33 bp    DNA     linear   PAT 16-APR-2003
DEFINITION  A novel perparation method for a object polypeptide by using UspA.
ACCESSION   BD179370
VERSION     BD179370.1  GI:30016640
KEYWORDS    WO 02083907-A/15.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 33)
  AUTHORS   Mitsuda,Y.
  TITLE     A novel perparation method for a object polypeptide by using UspA
  JOURNAL   Patent: WO 02083907-A 15 24-OCT-2002;
            SHIONOGI AND CO LTD,YUICHI MITSUDA
COMMENT     OS   Artificial Sequence
            PN   WO 02083907-A/15
            PD   24-OCT-2002
            PF   04-APR-2002 WO 2002JP003374
            PR   10-APR-2001 JP 01P   111088
            PI   YUICHI MITSUDA
            PC
            C12N15/62,C07K19/00,C12P21/02,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
             00
            CC   Description of Artificial Sequence:Sense primer for HdeB FH
            Key             Location/Qualifiers
            FT   source          1. .33
            FT                   /organism='Artificial Sequence'.
FEATURES              Location/Qualifiers
     source           1. .33
                      /organism="synthetic construct"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:32630"
BASE COUNT       10 a      8 c      4 g     11 t
ORIGIN

  Query Match            51.7%;  Score 15;  DB 6;  Length 33;
  Best Local Similarity   47.8%;  Pred. No. 4.4e+04;
  Matches   11;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 AAAGAUUCUUUUUGUAAGCCCCA 23
              ||| |::| ::::|:|| |  ||
Db         24 AAATATTCATTTTGTAACCTGCA 2


RESULT 10
AX514720/c
LOCUS       AX514720                  41 bp    DNA     linear   PAT 05-OCT-2002
DEFINITION  Sequence 918 from Patent WO02052044.
ACCESSION   AX514720
VERSION     AX514720.1  GI:23561343
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nakamura,Y., Sekine,A., Iida,A. and Saito,S.
  TITLE     Detection of genetic polymorphisms
  JOURNAL   Patent: WO 02052044-A 918 04-JUL-2002;
            Riken (JP)
FEATURES              Location/Qualifiers
     source           1. .41
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
BASE COUNT       12 a      10 c      10 g      8 t      1 others
ORIGIN

  Query Match            51.7%;  Score 15;  DB 6;  Length 41;
  Best Local Similarity   52.0%;  Pred. No. 4.3e+04;
  Matches   13;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGC 28
              ||::|  :::| ||||||  :|| |
Db         41 GATTCACTTTGCAAGCCCTCWGGAC 17


RESULT 11
AX520728/c
LOCUS       AX520728                  41 bp    DNA     linear   PAT 05-OCT-2002
DEFINITION  Sequence 6926 from Patent WO02052044.
ACCESSION   AX520728
VERSION     AX520728.1  GI:23571381
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nakamura,Y., Sekine,A., Iida,A. and Saito,S.
  TITLE     Detection of genetic polymorphisms
  JOURNAL   Patent: WO 02052044-A 6926 04-JUL-2002;
            Riken (JP)
FEATURES              Location/Qualifiers
     source           1. .41
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
BASE COUNT       12 a      10 c      10 g      8 t      1 others
ORIGIN

  Query Match            51.7%;  Score 15;  DB 6;  Length 41;
  Best Local Similarity   52.0%;  Pred. No. 4.3e+04;
  Matches   13;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGC 28
              ||::|  :::| ||||||  :|| |
Db         41 GATTCACTTTGCAAGCCCTCWGGAC 17


RESULT 12
AR019319/c
LOCUS       AR019319                  29 bp    DNA     linear   PAT 05-DEC-1998
DEFINITION  Sequence 25 from patent US 5783406.
ACCESSION   AR019319
VERSION     AR019319.1  GI:3974433
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
```

```
           Unclassified.
REFERENCE   1  (bases 1 to 29)
  AUTHORS   Schumm,J.W. and Puers,C.
  TITLE     Allelic ladders for short tandem repeat loci
  JOURNAL   Patent: US 5783406-A 25 21-JUL-1998;
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="unknown"
BASE COUNT       12 a      4 c      7 g      6 t
ORIGIN

  Query Match           51.0%;  Score 14.8;  DB 6;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 5.5e+04;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGCU 29
              ||:: :: :: : | || | |||||:
Db         29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT 13
AR061847/c
LOCUS       AR061847                 29 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 39 from patent US 5843660.
ACCESSION   AR061847
VERSION     AR061847.1  GI:5989538
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 29)
  AUTHORS   Schumm,J.W., Micka,K.A. and Rabbach,D.R.
  TITLE     Multiplex amplification of short tandem repeat loci
  JOURNAL   Patent: US 5843660-A 39 01-DEC-1998;
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="unknown"
BASE COUNT       12 a      4 c      7 g      6 t
ORIGIN

  Query Match           51.0%;  Score 14.8;  DB 6;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 5.5e+04;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGCU 29
              ||:: :: :: : | || | |||||:
Db         29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT 14
AR147578/c
LOCUS       AR147578                 29 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 31 from patent US 6221598.
ACCESSION   AR147578
VERSION     AR147578.1  GI:15111381
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 29)
  AUTHORS   Schumm,J.W., Sprecher,C.J. and Lins,A.M.
  TITLE     Multiplex amplification of short tandem repeat loci
  JOURNAL   Patent: US 6221598-A 31 24-APR-2001;
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="unknown"
BASE COUNT       12 a      4 c      7 g      6 t
ORIGIN

  Query Match           51.0%;  Score 14.8;  DB 6;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 5.5e+04;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGCU 29
              ||:: :: :: : | || | |||||:
Db         29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT 15
AR252838/c
LOCUS       AR252838                 29 bp    DNA     linear   PAT 20-DEC-2002
DEFINITION  Sequence 39 from patent US 6479235.
ACCESSION   AR252838
VERSION     AR252838.1  GI:27301187
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 29)
  AUTHORS   Schumm,J.W. and Sprecher,C.J.
  TITLE     Multiplex amplification of short tandem repeat loci
  JOURNAL   Patent: US 6479235-A 39 12-NOV-2002;
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="unknown"
BASE COUNT       12 a      4 c      7 g      6 t
ORIGIN

  Query Match           51.0%;  Score 14.8;  DB 6;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 5.5e+04;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGCU 29
              ||:: :: :: : | || | |||||:
Db         29 GATTATTCTTATCATCCACTAGGGCT 4


Search completed: January 30, 2004, 08:51:26
Job time : 578.333 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 06:19:17 ; Search time 283.333 Seconds
(without alignments)
276.295 Million cell updates/sec

Title: US-09-310-844C-25
Perfect score: 29
Sequence: 1 aaagauucuuuuuguaagccccaagggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 2640686

Minimum DB seq length: 0
Maximum DB seq length: 70

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	29	100.0	29	21	AAA70829	Molecular interact
2	29	100.0	29	21	AAA70830	Molecular interact
3	29	100.0	42	21	AAA71115	Molecular interact
4	29	100.0	42	21	AAA71116	Molecular interact
5	29	100.0	42	21	AAA71120	Molecular interact
6	29	100.0	42	21	AAA71121	Molecular interact
7	29	100.0	42	21	AAA71128	Molecular interact
8	29	100.0	42	21	AAA71129	Molecular interact
9	28	96.6	45	21	AAA70825	Molecular interact
10	28	96.6	45	21	AAA70826	Molecular interact
11	28	96.6	46	21	AAA71088	Molecular interact
12	28	96.6	46	21	AAA71089	Molecular interact
13	28	96.6	46	21	AAA71090	Molecular interact
14	28	96.6	46	21	AAA71105	Molecular interact
15	28	96.6	46	21	AAA71106	Molecular interact
16	28	96.6	46	21	AAA71107	Molecular interact
17	24.8	85.5	42	21	AAA71113	Molecular interact
18	24.8	85.5	42	21	AAA71118	Molecular interact
19	24.8	85.5	42	21	AAA71126	Molecular interact
20	23.8	82.1	46	21	AAA71085	Molecular interact
21	23.8	82.1	46	21	AAA71103	Molecular interact
22	23.2	80.0	29	21	AAA70828	Molecular interact
23	23.2	80.0	42	21	AAA71123	Molecular interact
24	23.2	80.0	42	21	AAA71131	Molecular interact
25	22.2	76.6	45	21	AAA70824	Molecular interact
26	22.2	76.6	46	21	AAA71087	Molecular interact
27	22.2	76.6	46	21	AAA71096	Molecular interact
28	22.2	76.6	46	21	AAA71099	Molecular interact
29	22.2	76.6	46	21	AAA71100	Molecular interact
30	22.2	76.6	46	21	AAA71104	Molecular interact
31	21.2	73.1	42	21	AAA71114	Molecular interact
32	21.2	73.1	42	21	AAA71119	Molecular interact
33	21.2	73.1	42	21	AAA71127	Molecular interact
34	21.2	73.1	46	21	AAA71094	Molecular interact
35	21.2	73.1	46	21	AAA71110	Molecular interact
36	20	69.0	46	21	AAA71084	Molecular interact
37	20	69.0	46	21	AAA71098	Molecular interact
38	20	69.0	46	21	AAA71102	Molecular interact
39	19.6	67.6	42	21	AAA71124	Molecular interact
40	19.6	67.6	42	21	AAA71132	Molecular interact
41	18.6	64.1	46	21	AAA71093	Molecular interact
42	18.6	64.1	46	21	AAA71095	Molecular interact
43	18.6	64.1	46	21	AAA71109	Molecular interact
44	18.6	64.1	46	21	AAA71111	Molecular interact
45	18.4	63.4	42	21	AAA71117	Molecular interact

ALIGNMENTS

RESULT 1
AAA70829
ID AAA70829 standard; RNA; 29 BP.
XX
AC AAA70829;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #29.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Mus sp.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,

PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Claim 235; Page 235; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 29 BP; 8 A; 6 C; 6 G; 9 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 29
||||||||||||||||||||||||||||||
Db 1 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 29

RESULT 2
AAA70830
ID AAA70830 standard; RNA; 29 BP.
XX
AC AAA70830;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #30.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Rattus sp.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Claim 235; Page 235; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 29 BP; 8 A; 6 C; 6 G; 9 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 29
||||||||||||||||||||||||||||||
Db 1 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 29

RESULT 3
AAA71115
ID AAA71115 standard; RNA; 42 BP.
XX
AC AAA71115;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #191.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -

XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 29
              ||||||||||||||||||||||||||||||
Db          4 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 32
```

RESULT 4
AAA71116
ID AAA71116 standard; RNA; 42 BP.
XX
AC AAA71116;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #192.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX

PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 29
||||||||||||||||||||||||||||||
Db 4 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 32

RESULT 5
AAA71120
ID AAA71120 standard; DNA; 42 BP.
XX
AC AAA71120;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #126.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 125; 405pp; English.

XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 T; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 69.0%; Pred. No. 0.0016;
Matches 20; Conservative 9; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 29
||||||::|:::::|:||||||||||||:
Db 4 AAAGATTCTTTTTGTAAGCCCCAAGGGCT 32

RESULT 6
AAA71121
ID AAA71121 standard; DNA; 42 BP.
XX
AC AAA71121;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #127.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 125; 405pp; English.
XX

CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 T; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 69.0%; Pred. No. 0.0016;
Matches 20; Conservative 9; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 29
||||||::|:::::|:||||||||||||:
Db 4 AAAGATTCTTTTTGTAAGCCCCAAGGGCT 32

RESULT 7
AAA71128
ID AAA71128 standard; RNA; 42 BP.
XX
AC AAA71128;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #197.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 126; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which

CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 29
||||||||||||||||||||||||||||
Db 4 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 32

RESULT 8
AAA71129
ID AAA71129 standard; RNA; 42 BP.
XX
AC AAA71129;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #198.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 126; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 29
||||||||||||||||||||||||||||
Db 4 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 32

RESULT 9
AAA70825
ID AAA70825 standard; RNA; 45 BP.
XX
AC AAA70825;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #25.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Mus sp.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Claim 221; Page 232; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of

CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 45 BP; 14 A; 7 C; 9 G; 15 U; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAGAUUCUUUUUGUAAGCCCCAAGGGC 28
||||||||||||||||||||||||||||
Db 18 AAAGAUUCUUUUUGUAAGCCCCAAGGGC 45

RESULT 10
AAA70826
ID AAA70826 standard; RNA; 45 BP.
XX
AC AAA70826;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #26.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Rattus sp.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Claim 222; Page 232; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 45 BP; 14 A; 7 C; 9 G; 15 U; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAGAUUCUUUUUGUAAGCCCCAAGGGC 28
||||||||||||||||||||||||||||
Db 18 AAAGAUUCUUUUUGUAAGCCCCAAGGGC 45

RESULT 11
AAA71088
ID AAA71088 standard; DNA; 46 BP.
XX
AC AAA71088;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #111.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 121; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual

CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 14 A; 7 C; 9 G; 16 T; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 71.4%; Pred. No. 0.0046;
Matches 20; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAGAUUCUUUUUGUAAGCCCCAAGGGC 28
||||||:|:::::|:||||||||||||
Db 19 AAAGATTCTTTTTGTAAGCCCCAAGGGC 46

RESULT 12
AAA71089
ID AAA71089 standard; DNA; 46 BP.
XX
AC AAA71089;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #112.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 121; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 14 A; 7 C; 9 G; 16 T; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 71.4%; Pred. No. 0.0046;
Matches 20; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAGAUUCUUUUUGUAAGCCCCAAGGGC 28
||||||:|:::::|:||||||||||||
Db 19 AAAGATTCTTTTTGTAAGCCCCAAGGGC 46

RESULT 13
AAA71090
ID AAA71090 standard; DNA; 46 BP.
XX
AC AAA71090;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #113.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 121; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)

CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 14 A; 7 C; 9 G; 16 T; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 71.4%; Pred. No. 0.0046;
Matches 20; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAAGAUUCUUUUUGUAAGCCCCAAGGGC 28
              |||||::|:::::|:||||||||||||||
Db         19 AAAGATTCTTTTTGTAAGCCCCAAGGGC 46
```

RESULT 14
AAA71105
ID AAA71105 standard; RNA; 46 BP.
XX
AC AAA71105;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #181.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 14 A; 7 C; 9 G; 16 U; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.0046;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAAGAUUCUUUUUGUAAGCCCCAAGGGC 28
              ||||||||||||||||||||||||||||
Db         19 AAAGAUUCUUUUUGUAAGCCCCAAGGGC 46
```

RESULT 15
AAA71106
ID AAA71106 standard; RNA; 46 BP.
XX
AC AAA71106;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #182.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance

CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 14 A; 7 C; 9 G; 16 U; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.0046;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAAGAUUCUUUUUGUAAGCCCCAAGGGC 28
              ||||||||||||||||||||||||||||
Db         19 AAAGAUUCUUUUUGUAAGCCCCAAGGGC 46
```

Search completed: January 30, 2004, 08:22:12
Job time : 283.667 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 07:56:58 ; Search time 50 Seconds
(without alignments)
256.002 Million cell updates/sec

Title: US-09-310-844C-25
Perfect score: 29
Sequence: 1 aaagauucuuuuuguaagccccaagggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 792150

Minimum DB seq length: 0
Maximum DB seq length: 70

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	15.2	52.4	33	1	US-08-667-079B-5	Sequence 5, Appli
	2	15.2	52.4	47	4	US-09-422-978-1097	Sequence 1097, Ap
c	3	14.8	51.0	29	1	US-08-219-633-25	Sequence 25, Appl
c	4	14.8	51.0	29	1	US-08-515-236-25	Sequence 25, Appl
c	5	14.8	51.0	29	1	US-08-761-950-25	Sequence 25, Appl
c	6	14.8	51.0	29	2	US-08-632-575B-39	Sequence 39, Appl
c	7	14.8	51.0	29	3	US-09-327-229-31	Sequence 31, Appl
c	8	14.8	51.0	29	4	US-09-199-542B-39	Sequence 39, Appl
c	9	14.8	51.0	29	5	PCT-US95-12608-31	Sequence 31, Appl
c	10	14.8	51.0	32	2	US-08-632-575B-59	Sequence 59, Appl
c	11	14.8	51.0	32	4	US-09-199-542B-59	Sequence 59, Appl
	12	14.6	50.3	25	4	US-09-063-733A-18	Sequence 18, Appl
c	13	14.6	50.3	53	2	US-08-486-969-46	Sequence 46, Appl
	14	14.2	49.0	47	4	US-09-422-978-96	Sequence 96, Appl
	15	14.2	49.0	69	2	US-08-410-654B-30	Sequence 30, Appl
	16	14.2	49.0	69	2	US-08-474-851-30	Sequence 30, Appl
	17	14.2	49.0	69	2	US-08-481-560-30	Sequence 30, Appl
	18	14	48.3	41	4	US-09-571-774-2	Sequence 2, Appli
c	19	13.8	47.6	25	3	US-08-943-731-336	Sequence 336, App
c	20	13.8	47.6	33	4	US-09-199-542B-76	Sequence 76, Appl
c	21	13.8	47.6	47	4	US-09-671-317-663	Sequence 663, App
	22	13.6	46.9	41	4	US-09-565-156A-2	Sequence 2, Appli
c	23	13.6	46.9	47	4	US-09-422-978-1843	Sequence 1843, Ap
c	24	13.6	46.9	47	4	US-09-402-266B-10	Sequence 10, Appl
	25	13.6	46.9	52	4	US-09-310-463-6	Sequence 6, Appli
	26	13.6	46.9	52	4	US-08-842-248A-6	Sequence 6, Appli
	27	13.4	46.2	32	3	US-08-718-738-16	Sequence 16, Appl
	28	13.4	46.2	32	3	US-09-221-844-16	Sequence 16, Appl
	29	13.4	46.2	32	5	PCT-US95-03323A-16	Sequence 16, Appl
c	30	13.4	46.2	46	1	US-08-171-389-42	Sequence 42, Appl
c	31	13.4	46.2	46	1	US-08-171-389-45	Sequence 45, Appl
c	32	13.4	46.2	46	1	US-08-123-936-42	Sequence 42, Appl
c	33	13.4	46.2	46	1	US-08-123-936-45	Sequence 45, Appl
c	34	13.4	46.2	46	2	US-08-475-228A-42	Sequence 42, Appl
c	35	13.4	46.2	46	2	US-08-475-228A-45	Sequence 45, Appl
c	36	13.4	46.2	46	3	US-08-482-080A-42	Sequence 42, Appl
c	37	13.4	46.2	46	3	US-08-482-080A-45	Sequence 45, Appl
c	38	13.4	46.2	46	4	US-09-354-947-42	Sequence 42, Appl
c	39	13.4	46.2	46	4	US-09-354-947-45	Sequence 45, Appl
c	40	13.4	46.2	46	5	PCT-US93-12388-42	Sequence 42, Appl
c	41	13.4	46.2	46	5	PCT-US93-12388-45	Sequence 45, Appl
	42	13.4	46.2	50	1	US-08-245-754A-13	Sequence 13, Appl
c	43	13.4	46.2	50	1	US-08-171-389-46	Sequence 46, Appl
c	44	13.4	46.2	50	1	US-08-123-936-46	Sequence 46, Appl
c	45	13.4	46.2	50	2	US-08-475-228A-46	Sequence 46, Appl

ALIGNMENTS

```
RESULT 1
US-08-667-079B-5/c
; Sequence 5, Application US/08667079B
; Patent No. 5789171
;  GENERAL INFORMATION:
;    APPLICANT:  Mark S. Smeltzer
;    TITLE OF INVENTION:  Use of cna, fnbA, fnbB, and hlb Gene Probes for the Strain-Sp
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Benjamin Aaron Adler, MCGREGOR & ADLER, P.C.
;      STREET:  8011 Candle Lane
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word for Macintosh
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/667,079B
;      FILING DATE:  June 20, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adler, Benjamin Aaron
;      REGISTRATION NUMBER:  35,423
;      REFERENCE/DOCKET NUMBER:  D5886
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-777-2321
;      TELEFAX:  713-777-6908
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  other nucleic acid
;    HYPOTHETICAL:  No
;    ANTI-SENSE:  No
;    ORIGINAL SOURCE:
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:
US-08-667-079B-5
```

```
  Query Match             52.4%;  Score 15.2;  DB 1;  Length 33;
  Best Local Similarity   42.9%;  Pred. No. 3.6e+02;
  Matches   12;  Conservative    8;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          2 AAGAUUCUUUUUGUAAGCCCCAAGGGCU 29
              | ||:: :::: |:||   ||  ||||:
Db         32 ATGATTGTTTTAGTAATTTCCCCGGGCT 5


RESULT 2
US-09-422-978-1097
; Sequence 1097, Application US/09422978
; Patent No. 6537751
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT: Chumakov, Ilya
;  TITLE OF INVENTION: Biallelic markers for use in constructing a high density...
;  FILE REFERENCE: GENSET.020CP1
;  CURRENT APPLICATION NUMBER: US/09/422,978
;  CURRENT FILING DATE: 1999-10-20
;  EARLIER APPLICATION NUMBER: US 09/298,850
;  EARLIER FILING DATE: 1999-04-21
;  EARLIER APPLICATION NUMBER: US 60/109,732
;  EARLIER FILING DATE: 1998-11-23
;  EARLIER APPLICATION NUMBER: US 60/082,614
;  EARLIER FILING DATE: 1998-04-21
;  NUMBER OF SEQ ID NOS: 11796
; SEQ ID NO 1097
;   LENGTH: 47
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 24
;   OTHER INFORMATION: 99-2043-220 : polymorphic base A or T
US-09-422-978-1097

  Query Match             52.4%;  Score 15.2;  DB 4;  Length 47;
  Best Local Similarity   50.0%;  Pred. No. 3.8e+02;
  Matches   11;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          3 AGAUUCUUUUUGUAAGCCCCAA 24
              ||| :|:::::|:|| | || |
Db         15 AGACTCTTTWTGTAACCTCCCA 36


RESULT 3
US-08-219-633-25/c
; Sequence 25, Application US/08219633
; Patent No. 5599666
;  GENERAL INFORMATION:
;    APPLICANT:  Schumm, James W.
;    APPLICANT:  Puers, Christoph
;    TITLE OF INVENTION:  ALLELIC LADDERS FOR SHORT TANDEM REPEAT
;    TITLE OF INVENTION:  LOCI
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ross & Stevens, S.C.
;      STREET:  One South Pinckney Street, P.O. Box 2599
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  U.S.A.
;      ZIP:  53701-2599
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/219,633
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sara, Charles S.
;      REGISTRATION NUMBER:  30,492
;      REFERENCE/DOCKET NUMBER:  34506.019
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 257-5353
;      TELEFAX:  (608) 257-9175
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-219-633-25

  Query Match             51.0%;  Score 14.8;  DB 1;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 5.3e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGCU 29
              |.|:: :: :: : | || | |||||:
Db         29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT 4
US-08-515-236-25/c
; Sequence 25, Application US/08515236
; Patent No. 5674686
;  GENERAL INFORMATION:
;    APPLICANT:  Schumm, James W.
;    APPLICANT:  Puers, Christoph
;    TITLE OF INVENTION:  ALLELIC LADDERS FOR SHORT TANDEM REPEAT
;    TITLE OF INVENTION:  LOCI
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ross & Stevens, S.C.
;      STREET:  One South Pinckney Street, P.O. Box 2599
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  U.S.A.
;      ZIP:  53701-2599
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/515,236
;      FILING DATE:  15-AUG-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/219,633
;      FILING DATE:  28-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sara, Charles S.
;      REGISTRATION NUMBER:  30,492
;      REFERENCE/DOCKET NUMBER:  34506.019
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 257-5353
;      TELEFAX:  (608) 257-9175
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-515-236-25

  Query Match             51.0%;  Score 14.8;  DB 1;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 5.3e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGCU 29
```

```
              ||:: :: :: : | || | |||||:
Db         29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT 5
US-08-761-950-25/c
; Sequence 25, Application US/08761950
; Patent No. 5783406
;  GENERAL INFORMATION:
;    APPLICANT:  Schumm, James W.
;    APPLICANT:  Puers, Christoph
;    TITLE OF INVENTION:  ALLELIC LADDERS FOR SHORT TANDEM REPEAT
;    TITLE OF INVENTION:  LOCI
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ross & Stevens, S.C.
;      STREET:  One South Pinckney Street, P.O. Box 2599
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  U.S.A.
;      ZIP:  53701-2599
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/761,950
;      FILING DATE:  09-DEC-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/515,236
;      FILING DATE:  15-AUG-1995
;      APPLICATION NUMBER:  US 08/219,633
;      FILING DATE:  28-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sara, Charles S.
;      REGISTRATION NUMBER:  30,492
;      REFERENCE/DOCKET NUMBER:  34506.019
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 257-5353
;      TELEFAX:  (608) 257-9175
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-761-950-25

  Query Match             51.0%;  Score 14.8;  DB 1;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 5.3e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGCU 29
              ||:: :: :: : | || | |||||:
Db         29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT 6
US-08-632-575B-39/c
; Sequence 39, Application US/08632575B
; Patent No. 5843660
;  GENERAL INFORMATION:
;    APPLICANT:  Schumm, James W.
;    TITLE OF INVENTION:  Multiplex Amplification of
;    TITLE OF INVENTION:  Short Tandem Repeat Loci
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Promega Corporation
;      STREET:  2800 Woods Hollow Road
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  U.S.A.
;      ZIP:  53711-5399
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette - 3.5 inch, 1.44 Mb
;      COMPUTER:  IBM compatible PC
;      OPERATING SYSTEM:  DOS, version 6.0
;      SOFTWARE:  WordPerfect 5.1 (DOS text format)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/632,575B
;      FILING DATE:  04/15/96
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/316,544
;      FILING DATE:  09/30/94
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    POSITION IN GENOME:
;      MAP POSITION:  HUMvWFA31
US-08-632-575B-39

  Query Match             51.0%;  Score 14.8;  DB 2;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 5.3e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGCU 29
              ||:: :: :: : | || | |||||:
Db         29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT 7
US-09-327-229-31/c
; Sequence 31, Application US/09327229
; Patent No. 6221598
;  GENERAL INFORMATION:
;       APPLICANT: Schumm, James W.
;                  Sprecher, Cynthia J.
;                  Lins, Ann M.
;       TITLE OF INVENTION: MULTIPLEX AMPLIFICATION OF SHORT TANDEM
;                           REPEAT LOCI
;       NUMBER OF SEQUENCES: 32
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Ross & Stevens, S.C.
;            STREET: P. O. Box 2599
;            CITY: Madison
;            STATE: Wisconsin
;            COUNTRY: U.S.A.
;            ZIP: 53701-2599
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/327,229
;            FILING DATE: 07-Jun-1999
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/316,544
;            FILING DATE: 30-SEP-1994
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Sara, Charles S.
;            REGISTRATION NUMBER: 30,492
;            REFERENCE/DOCKET NUMBER: 34506.022
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 608-257-5353
;            TELEFAX: 608-257-9175
;  INFORMATION FOR SEQ ID NO: 31:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 29 base pairs
;            TYPE: nucleic acid
```

```
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-327-229-31

  Query Match             51.0%;  Score 14.8;  DB 3;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 5.3e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGCU 29
              ||:: :: :: : | || | |||||:
Db         29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT 8
US-09-199-542B-39/c
; Sequence 39, Application US/09199542B
; Patent No. 6479235
; GENERAL INFORMATION:
;  APPLICANT:  Schumm, James W.
;  APPLICANT:  Sprecher, Cynthia J.
;  TITLE OF INVENTION:  Multiplex Amplification of Short Tandem Repeat Loci
;  FILE REFERENCE:  16026/9212
;  CURRENT APPLICATION NUMBER: US/09/199,542B
;  CURRENT FILING DATE:  1998-11-25
;  PRIOR APPLICATION NUMBER:  US 08/316,544
;  PRIOR FILING DATE:  1994-09-30
;  PRIOR APPLICATION NUMBER:  US 08/632,575
;  PRIOR FILING DATE:  1996-04-15
;  NUMBER OF SEQ ID NOS:  110
;  SOFTWARE:  Word97 (converted to DOS text format)
; SEQ ID NO 39
;   LENGTH: 29
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   LOCATION: HUMvWFA31
US-09-199-542B-39

  Query Match             51.0%;  Score 14.8;  DB 4;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 5.3e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGCU 29
              ||:: :: :: : | || | |||||:
Db         29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT 9
PCT-US95-12608-31/c
; Sequence 31, Application PC/TUS9512608
;  GENERAL INFORMATION:
;    APPLICANT:  Schumm, James W.
;    APPLICANT:  Sprecher, Cynthia J.
;    APPLICANT:  Lins, Ann M.
;    TITLE OF INVENTION:  MULTIPLEX AMPLIFICATION OF SHORT TANDEM
;    TITLE OF INVENTION:  REPEAT LOCI
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ross & Stevens, S.C.
;      STREET:  P. O. Box 2599
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  U.S.A.
;      ZIP:  53701-2599
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/12608
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sara, Charles S.
;      REGISTRATION NUMBER:  30,492
;      REFERENCE/DOCKET NUMBER:  34506.022
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-257-5353
;      TELEFAX:  608-257-9175
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US95-12608-31

  Query Match             51.0%;  Score 14.8;  DB 5;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 5.3e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGCU 29
              ||:: :: :: : | || | |||||:
Db         29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT 10
US-08-632-575B-59/c
; Sequence 59, Application US/08632575B
; Patent No. 5843660
;  GENERAL INFORMATION:
;    APPLICANT:  Schumm, James W.
;    TITLE OF INVENTION:  Multiplex Amplification of
;    TITLE OF INVENTION:  Short Tandem Repeat Loci
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Promega Corporation
;      STREET:  2800 Woods Hollow Road
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  U.S.A.
;      ZIP:  53711-5399
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette - 3.5 inch, 1.44 Mb
;      COMPUTER:  IBM compatible PC
;      OPERATING SYSTEM:  DOS, version 6.0
;      SOFTWARE:  WordPerfect 5.1 (DOS text format)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/632,575B
;      FILING DATE:  04/15/96
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/316,544
;      FILING DATE:  09/30/94
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    POSITION IN GENOME:
;      MAP POSITION:  HUMvWFA31
US-08-632-575B-59

  Query Match             51.0%;  Score 14.8;  DB 2;  Length 32;
  Best Local Similarity   42.3%;  Pred. No. 5.4e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGCU 29
              ||:: :: :: : | || | |||||:
Db         29 GATTATTCTTATCATCCACTAGGGCT 4
```

```
RESULT 11
US-09-199-542B-59/c
; Sequence 59, Application US/09199542B
; Patent No. 6479235
; GENERAL INFORMATION:
;  APPLICANT:  Schumm, James W.
;  APPLICANT:  Sprecher, Cynthia J.
;  TITLE OF INVENTION:  Multiplex Amplification of Short Tandem Repeat Loci
;  FILE REFERENCE:  16026/9212
;  CURRENT APPLICATION NUMBER: US/09/199,542B
;  CURRENT FILING DATE:  1998-11-25
;  PRIOR APPLICATION NUMBER:  US 08/316,544
;  PRIOR FILING DATE:  1994-09-30
;  PRIOR APPLICATION NUMBER:  US 08/632,575
;  PRIOR FILING DATE:  1996-04-15
;  NUMBER OF SEQ ID NOS:  110
;  SOFTWARE:  Word97 (converted to DOS text format)
; SEQ ID NO 59
;   LENGTH: 32
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   LOCATION: HUMvWFA31
US-09-199-542B-59

  Query Match             51.0%;  Score 14.8;  DB 4;  Length 32;
  Best Local Similarity   42.3%;  Pred. No. 5.4e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGCU 29
              ||:: :: :: : | || | |||||:
Db         29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT 12
US-09-063-733A-18
; Sequence 18, Application US/09063733A
; Patent No. 6372211
;  GENERAL INFORMATION:
;    APPLICANT:  Isaac, Barbara G.
;    APPLICANT:  Greenplate, John T.
;    APPLICANT:  Purcell, John P.
;    APPLICANT:  Romano, Charles P.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR CONTROLLING
;    TITLE OF INVENTION:  INSECTS
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold White & Durkee
;      STREET:  PO Box 4433
;      CITY:  Houston
;      STATE:  TX
;      COUNTRY:  USA
;      ZIP:  77210-4433
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/063,733A
;      FILING DATE:  21-APR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Patterson, Melinda L.
;      REGISTRATION NUMBER:  33,062
;      REFERENCE/DOCKET NUMBER:  MOBT:022
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-787-1400
;      TELEFAX:  713-787-1440
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  25 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-063-733A-18

  Query Match             50.3%;  Score 14.6;  DB 4;  Length 25;
  Best Local Similarity   47.6%;  Pred. No. 6.4e+02;
  Matches   10;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 AAGAUUCUUUUUGUAAGCCCC 22
              ||| ::|: :::|:||  |||
Db          5 AAGCTTCTCTTTGTAATACCC 25


RESULT 13
US-08-486-969-46/c
; Sequence 46, Application US/08486969
; Patent No. 5843456
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Maki, Joanne
;    TITLE OF INVENTION:  RECOMBINANT POXVIRUS - RABIES
;    TITLE OF INVENTION:  COMPOSITIONS AND COMBINATION COMPOSITIONS AND USES
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue, 25th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/486,969
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2600
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-486-969-46

  Query Match             50.3%;  Score 14.6;  DB 2;  Length 53;
  Best Local Similarity   47.6%;  Pred. No. 7.2e+02;
  Matches   10;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          9 UUUUUGUAAGCCCCAAGGGCU 29
              :::::|:|||| |  ||||:
Db         29 TTTTTGTAAGCTTCCCGGGCT 9


RESULT 14
US-09-422-978-96
; Sequence 96, Application US/09422978
; Patent No. 6537751
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT: Chumakov, Ilya
;  TITLE OF INVENTION: Biallelic markers for use in constructing a high density...
;  FILE REFERENCE: GENSET.020CP1
```

```
;  CURRENT APPLICATION NUMBER: US/09/422,978
;  CURRENT FILING DATE: 1999-10-20
;  EARLIER APPLICATION NUMBER: US 09/298,850
;  EARLIER FILING DATE: 1999-04-21
;  EARLIER APPLICATION NUMBER: US 60/109,732
;  EARLIER FILING DATE: 1998-11-23
;  EARLIER APPLICATION NUMBER: US 60/082,614
;  EARLIER FILING DATE: 1998-04-21
;  NUMBER OF SEQ ID NOS: 11796
; SEQ ID NO 96
;   LENGTH: 47
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 24
;   OTHER INFORMATION: 99-12847-37 : polymorphic base A or G
US-09-422-978-96

  Query Match             49.0%;  Score 14.2;  DB 4;  Length 47;
  Best Local Similarity   42.9%;  Pred. No. 1.1e+03;
  Matches    9;  Conservative    8;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          9 UUUUUGUAAGCCCCAAGGGCU 29
              ::::: :|||  ||| :|||:
Db          8 TTTTTCTAAGGTCCACRGGCT 28


RESULT 15
US-08-410-654B-30
; Sequence 30, Application US/08410654B
; Patent No. 5833976
;  GENERAL INFORMATION:
;    APPLICANT:  Rene de Waal Malefyt
;    APPLICANT:  Di-Hwei Hsu
;    APPLICANT:  Anne O'Garra
;    APPLICANT:  Hergen Spits
;    TITLE OF INVENTION:  Use of Interleukin-10 to Treat
;    TITLE OF INVENTION:  Septic Shock
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering-Plough Corporation
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  7.5.3
;      SOFTWARE:  Microsoft Word 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/410,654B
;      FILING DATE:  24-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/229,854
;      FILING DATE:  19-APR-1994
;      APPLICATION NUMBER:  US 07/926,853
;      FILING DATE:  06-AUG-1992
;      APPLICATION NUMBER:  US 07/742,129
;      FILING DATE:  06-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Foulke, Cynthia L.
;      REGISTRATION NUMBER:  32,364
;      REFERENCE/DOCKET NUMBER:  DX0221KQ1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  908-298-2987
;      TELEFAX:  908-298-5388
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  69 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (oligonucleotide)
US-08-410-654B-30

  Query Match             49.0%;  Score 14.2;  DB 2;  Length 69;
  Best Local Similarity   51.9%;  Pred. No. 1.2e+03;
  Matches   14;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 AAAGAUUCUUUUUGUAAGCCCCAAGGG 27
              ||  |: | :::  :|||| ||||| |
Db          7 AAGAATGCCTTTAATAAGCTCCAAGAG 33


Search completed: January 30, 2004, 10:15:14
Job time : 51 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 08:51:33 ; Search time 176 Seconds
(without alignments)
600.524 Million cell updates/sec

Title: US-09-310-844C-25
Perfect score: 29
Sequence: 1 aaagauucuuuuuguaagccccaagggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 1462844

Minimum DB seq length: 0
Maximum DB seq length: 70

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	15.4	53.1	53	10	US-09-983-965-4754	Sequence 4754, Ap
	2	15.2	52.4	47	12	US-10-349-143-1097	Sequence 1097, Ap
c	3	15.2	52.4	60	13	US-09-908-975-18725	Sequence 18725, A
	4	15	51.7	39	15	US-10-116-519-18	Sequence 18, Appl
c	5	15	51.7	60	13	US-09-908-975-12187	Sequence 12187, A
	6	15	51.7	65	13	US-09-908-975-4580	Sequence 4580, Ap
c	7	14.8	51.0	29	13	US-09-839-478-31	Sequence 31, Appl
	8	14.6	50.3	25	15	US-10-005-530-18	Sequence 18, Appl
	9	14.6	50.3	60	13	US-09-908-975-8435	Sequence 8435, Ap
	10	14.6	50.3	60	13	US-09-908-975-18114	Sequence 18114, A
	11	14.6	50.3	60	13	US-10-378-094-45	Sequence 45, Appl
c	12	14.4	49.7	31	11	US-09-848-754A-6937	Sequence 6937, Ap
c	13	14.4	49.7	31	11	US-09-740-332-5660	Sequence 5660, Ap
c	14	14.4	49.7	31	13	US-09-817-879-5660	Sequence 5660, Ap
c	15	14.4	49.7	65	13	US-09-908-975-3924	Sequence 3924, Ap
	16	14.4	49.7	65	13	US-09-908-975-30555	Sequence 30555, A
c	17	14.2	49.0	25	10	US-09-827-998-1098	Sequence 1098, Ap
c	18	14.2	49.0	25	10	US-09-827-998-1099	Sequence 1099, Ap
c	19	14.2	49.0	25	10	US-09-827-998-1100	Sequence 1100, Ap
c	20	14.2	49.0	25	10	US-09-827-998-1101	Sequence 1101, Ap
c	21	14.2	49.0	25	10	US-09-827-998-1102	Sequence 1102, Ap
c	22	14.2	49.0	25	10	US-09-827-998-1103	Sequence 1103, Ap
c	23	14.2	49.0	25	10	US-09-827-998-1104	Sequence 1104, Ap
	24	14.2	49.0	25	15	US-10-098-263B-76444	Sequence 76444, A
	25	14.2	49.0	26	15	US-10-287-919-545	Sequence 545, App
c	26	14.2	49.0	31	11	US-09-848-754A-6794	Sequence 6794, Ap
c	27	14.2	49.0	31	11	US-09-740-332-7606	Sequence 7606, Ap
c	28	14.2	49.0	31	13	US-09-817-879-7606	Sequence 7606, Ap
	29	14.2	49.0	47	12	US-10-349-143-96	Sequence 96, Appl
c	30	14.2	49.0	50	12	US-10-131-827-2080	Sequence 2080, Ap
	31	14.2	49.0	60	13	US-09-908-975-13840	Sequence 13840, A
c	32	14.2	49.0	60	13	US-09-908-975-13940	Sequence 13940, A
c	33	14.2	49.0	60	13	US-09-908-975-16808	Sequence 16808, A
	34	14.2	49.0	65	10	US-09-783-590-1151	Sequence 1151, Ap
	35	14.2	49.0	65	13	US-10-032-585-500	Sequence 500, App
	36	14	48.3	41	11	US-09-852-385-2	Sequence 2, Appli
	37	14	48.3	57	11	US-09-860-738C-46	Sequence 46, Appl
c	38	13.8	47.6	27	9	US-09-823-936-33	Sequence 33, Appl
c	39	13.8	47.6	27	14	US-10-193-142-10	Sequence 10, Appl
c	40	13.8	47.6	27	15	US-10-134-493-10	Sequence 10, Appl
c	41	13.8	47.6	27	15	US-10-061-216-10	Sequence 10, Appl
c	42	13.8	47.6	27	15	US-10-286-140-33	Sequence 33, Appl
c	43	13.8	47.6	31	11	US-09-740-332-9154	Sequence 9154, Ap
c	44	13.8	47.6	31	13	US-09-817-879-9154	Sequence 9154, Ap
	45	13.8	47.6	55	13	US-10-222-654-96	Sequence 96, Appl

ALIGNMENTS

RESULT 1
US-09-983-965-4754
; Sequence 4754, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4754
; LENGTH: 53
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 18-LIB34-011-Q1-E1-E5
US-09-983-965-4754

Query Match 53.1%; Score 15.4; DB 10; Length 53;
Best Local Similarity 44.0%; Pred. No. 3e+03;
Matches 11; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

```
Qy          5 AUUCUUUUUGUAAGCCCCAAGGGCU 29
              |::|::: :|:  |||   |||||:
Db         28 ATTCTTTGTGTTTGCCTTCAGGGCT 52
```

RESULT 2
US-10-349-143-1097

```
; Sequence 1097, Application US/10349143
; Publication No. US20040005584A1
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  TITLE OF INVENTION: Biallelic markers for use in constructing a high density...
;  FILE REFERENCE: GENSET.020CP1
;  CURRENT APPLICATION NUMBER: US/10/349,143
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: US/09/422,978
;  PRIOR FILING DATE: 1999-10-20
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 09/298,850
;  PRIOR FILING DATE: EARLIER FILING DATE: 1999-04-21
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/109,732
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-23
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/082,614
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-21
;  NUMBER OF SEQ ID NOS: 11796
; SEQ ID NO 1097
;   LENGTH: 47
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 24
;   OTHER INFORMATION: 99-2043-220 : polymorphic base A or T
US-10-349-143-1097

  Query Match             52.4%;  Score 15.2;  DB 12;  Length 47;
  Best Local Similarity   50.0%;  Pred. No. 3.6e+03;
  Matches   11;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          3 AGAUUCUUUUUGUAAGCCCCAA 24
              ||| :|:::::|:|| | || |
Db         15 AGACTCTTTWTGTAACCTCCCA 36


RESULT 3
US-09-908-975-18725/c
; Sequence 18725, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18725
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-908-975-18725

  Query Match             52.4%;  Score 15.2;  DB 13;  Length 60;
  Best Local Similarity   60.0%;  Pred. No. 3.7e+03;
  Matches   12;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          7 UCUUUUUGUAAGCCCCAAGG 26
              :|::: :| |||||||| ||
Db         26 TCTTTCTGAAAGCCCCATGG 7


RESULT 4
US-10-116-519-18
; Sequence 18, Application US/10116519
; Publication No. US20030114373A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL CYSTEINE PROTEASE OF THE CALPAIN
;  TITLE OF INVENTION:  SUPERFAMILY, CAN-12 AND VARIANTS THEREOF
;  FILE REFERENCE: D0124 NP
;  CURRENT APPLICATION NUMBER: US/10/116,519
;  CURRENT FILING DATE:  2002-04-03
;  PRIOR APPLICATION NUMBER: US 60/281,253
;  PRIOR FILING DATE: 2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/288,768
;  PRIOR FILING DATE: 2001-05-04
;  PRIOR APPLICATION NUMBER: US 60/296,180
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/300,620
;  PRIOR FILING DATE: 2001-06-25
;  NUMBER OF SEQ ID NOS: 145
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
;   LENGTH: 39
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-116-519-18

  Query Match             51.7%;  Score 15;  DB 15;  Length 39;
  Best Local Similarity   56.5%;  Pred. No. 4.2e+03;
  Matches   13;  Conservative    5;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGG 26
              | ::| :: :| |||| ||||||
Db          5 GGTTCCTTCTGAAAGCTCCAAGG 27


RESULT 5
US-09-908-975-12187/c
; Sequence 12187, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE \
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12187
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-908-975-12187

  Query Match             51.7%;  Score 15;  DB 13;  Length 60;
  Best Local Similarity   47.8%;  Pred. No. 4.6e+03;
  Matches   11;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGAUUCUUUUUGUAAGCCCCAAG 25
              |||::|::: :|:|  | | |||
Db         55 AGATTCTTTCTGTAGCCGCTAAG 33
```

```
RESULT 6
US-09-908-975-4580
; Sequence 4580, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4580
;   LENGTH: 65
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-09-908-975-4580

  Query Match             51.7%;  Score 15;  DB 13;  Length 65;
  Best Local Similarity   56.5%;  Pred. No. 4.7e+03;
  Matches   13;  Conservative    5;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 AAGAUUCUUUUUGUAAGCCCCAA 24
              ||||: |:: ::| ||||  |||
Db         34 AAGATGCTTCTTGAAAGCAACAA 56


RESULT 7
US-09-839-478-31/c
; Sequence 31, Application US/09839478
; Publication No. US20030180724A1
;   GENERAL INFORMATION:
;        APPLICANT: Schumm, James W.
;                   Sprecher, Cynthia J.
;                   Lins, Ann M.
;        TITLE OF INVENTION: MULTIPLEX AMPLIFICATION OF SHORT TANDEM
;                            REPEAT LOCI
;        NUMBER OF SEQUENCES: 32
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Ross & Stevens, S.C.
;             STREET: P. O. Box 2599
;             CITY: Madison
;             STATE: Wisconsin
;             COUNTRY: U.S.A.
;             ZIP: 53701-2599
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/839,478
;             FILING DATE: 20-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/316,544
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Sara, Charles S.
;             REGISTRATION NUMBER: 30,492
;             REFERENCE/DOCKET NUMBER: 34506.022
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 608-257-5353
;             TELEFAX: 608-257-9175
;   INFORMATION FOR SEQ ID NO: 31:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 29 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-839-478-31

  Query Match             51.0%;  Score 14.8;  DB 13;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 4.9e+03;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGCU 29
              ||:: :: :: : | || | |||||:
Db         29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT 8
US-10-005-530-18
; Sequence 18, Application US/10005530
; Publication No. US20030026795A1
; GENERAL INFORMATION:
;  APPLICANT: Isaac, Barbara G.
;  APPLICANT:  Greenplate, John T.
;  APPLICANT:  Purcell, John P.
;  APPLICANT:  Romano, Charles P.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING INSECTS
;  FILE REFERENCE: 11899.0022.DVUS01 (MOBT:022--2)
;  CURRENT APPLICATION NUMBER: US/10/005,530
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: 09/063,733
;  PRIOR FILING DATE: 1998-04-21
;  PRIOR APPLICATION NUMBER: 60/044,504
;  PRIOR FILING DATE: 1997-04-21
;  NUMBER OF SEQ ID NOS: 58
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
;   LENGTH: 25
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Oligonucleotide
US-10-005-530-18

  Query Match             50.3%;  Score 14.6;  DB 15;  Length 25;
  Best Local Similarity   47.6%;  Pred. No. 5.9e+03;
  Matches   10;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 AAGAUUCUUUUUGUAAGCCCC 22
              ||| ::|: :::|:||  |||
Db          5 AAGCTTCTCTTTGTAATACCC 25


RESULT 9
US-09-908-975-8435
; Sequence 8435, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE \
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
```

```
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8435
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-908-975-8435

  Query Match             50.3%;  Score 14.6;  DB 13;  Length 60;
  Best Local Similarity   51.7%;  Pred. No. 7e+03;
  Matches   15;  Conservative    5;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 AAAGAUUCUUUUUGUAAGCCCCAAGGGCU 29
              || ||  |:  ::|:|| ||||| |  |:
Db         12 AACGAAACTGATTGTAATCCCCAGGATCT 40


RESULT 10
US-09-908-975-18114
; Sequence 18114, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18114
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-908-975-18114

  Query Match             50.3%;  Score 14.6;  DB 13;  Length 60;
  Best Local Similarity   52.4%;  Pred. No. 7e+03;
  Matches   11;  Conservative    6;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          9 UUUUUGUAAGCCCCAAGGGCU 29
              : ::: : || |||||||||:
Db          7 TATTTCTGAGTCCCAAGGGCT 27


RESULT 11
US-10-378-094-45
; Sequence 45, Application US/10378094
; Publication No. US20030221201A1
; GENERAL INFORMATION:
;  APPLICANT: PRIOR, Christopher P.
;  APPLICANT:  LAI, Char-Huei
;  APPLICANT:  SADEGHI, Homayoun
;  APPLICANT:  TURNER, Andrew
;  TITLE OF INVENTION: MODIFIED TRANSFERRIN FUSION PROTEINS
;  FILE REFERENCE: 54710-5001-01-US
;  CURRENT APPLICATION NUMBER: US/10/378,094
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: US 10/231,494
;  PRIOR FILING DATE: 2002-08-30
;  PRIOR APPLICATION NUMBER: US 60/334,059
;  PRIOR FILING DATE: 2001-11-30
;  PRIOR APPLICATION NUMBER: US 60/315,745
;  PRIOR FILING DATE: 2001-08-30
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 45
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic oligonucleotide encoding peptide with EPO activity
US-10-378-094-45

  Query Match             50.3%;  Score 14.6;  DB 13;  Length 60;
  Best Local Similarity   52.4%;  Pred. No. 7e+03;
  Matches   11;  Conservative    6;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          6 UUCUUUUUGUAAGCCCCAAGG 26
              ::   :::|:||||| |||||
Db         36 TTGGGTTTGTAAGCCACAAGG 56


RESULT 12
US-09-848-754A-6937/c
; Sequence 6937, Application US/09848754A
; Publication No. US20030073207A1
; GENERAL INFORMATION:
;  APPLICANT: Ribozyme Pharmaceuticals, Inc.
;  TITLE OF INVENTION: Enzymatic Nucleic Acid Treatment of Diseases or Conditions Relatec
;  TITLE OF INVENTION:  Levels of Epidermal Growth Factor Receptors
;  FILE REFERENCE: MBHB00-958-I (400/018)
;  CURRENT APPLICATION NUMBER: US/09/848,754A
;  CURRENT FILING DATE:  2001-05-03
;  NUMBER OF SEQ ID NOS: 9645
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6937
;   LENGTH: 31
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Enzymatic Nucleic acid
US-09-848-754A-6937

  Query Match             49.7%;  Score 14.4;  DB 11;  Length 31;
  Best Local Similarity   58.3%;  Pred. No. 7.6e+03;
  Matches   14;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGG 27
              ||:  :: : |  ||||||||||
Db         25 GATCGTTGTAGCTAGCCCCAAGGG 2


RESULT 13
US-09-740-332-5660/c
; Sequence 5660, Application US/09740332
; Publication No. US20030125270A1
; GENERAL INFORMATION:
;  APPLICANT: Ribozyme Pharmaceuticals Inc.
;  TITLE OF INVENTION: Enzymatic Nucleic Acid Treatment of Diseases or Conditions Relatec
;  TITLE OF INVENTION:  Hepatitis C Virus Infection
;  FILE REFERENCE: RPI 400/003
;  CURRENT APPLICATION NUMBER: US/09/740,332
;  CURRENT FILING DATE:  2001-03-26
;  NUMBER OF SEQ ID NOS: 9704
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5660
;   LENGTH: 31
;   TYPE: DNA
;   ORGANISM: artificial sequence
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION:
;   OTHER INFORMATION: DNAzyme
US-09-740-332-5660

  Query Match             49.7%;  Score 14.4;  DB 11;  Length 31;
```

Best Local Similarity 58.3%; Pred. No. 7.6e+03;
Matches 14; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

```
Qy          2 AAGAUUCUUUUUGUAAGCCCCAAG 25
              ||||:  :: : |  |||||||||
Db         27 AAGATCGTTGTAGCTAGCCCCAAG 4
```

RESULT 14
US-09-817-879-5660/c
; Sequence 5660, Application US/09817879
; Publication No. US20030171311A1
; GENERAL INFORMATION:
; APPLICANT: Ribozyme Pharmaceuticals Inc.
; TITLE OF INVENTION: Enzymatic Nucleic Acid Treatment of Diseases or Conditions Relate
; TITLE OF INVENTION: Hepatitis C Virus Infection
; FILE REFERENCE: MBHB00-801-F
; CURRENT APPLICATION NUMBER: US/09/817,879
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 9703
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5660
; LENGTH: 31
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: DNAzyme
US-09-817-879-5660

Query Match 49.7%; Score 14.4; DB 13; Length 31;
Best Local Similarity 58.3%; Pred. No. 7.6e+03;
Matches 14; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

```
Qy          2 AAGAUUCUUUUUGUAAGCCCCAAG 25
              ||||:  :: : |  |||||||||
Db         27 AAGATCGTTGTAGCTAGCCCCAAG 4
```

RESULT 15
US-09-908-975-3924/c
; Sequence 3924, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3924
; LENGTH: 65
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-908-975-3924

Query Match 49.7%; Score 14.4; DB 13; Length 65;
Best Local Similarity 50.0%; Pred. No. 8.8e+03;
Matches 12; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

```
Qy          1 AAAGAUUCUUUUUGUAAGCCCCAA 24
              | |||::|  :::|:||| |  ||
Db         35 AGAGATTCCCTTTGTAAGGCGTAA 12
```

Search completed: January 30, 2004, 13:10:26
Job time : 178 secs

GenCore version 5.1.6

QM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 07:55:03 ; Search time 1615 Seconds
(without alignments)
436.427 Million cell updates/sec

Title: US-09-310-844C-25
Perfect score: 29
Sequence: 1 aaagauucuuuuuguaagccccaagggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 243536

Minimum DB seq length: 0
Maximum DB seq length: 70

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	18.4	63.4	70	9	AA516989	AA516989 vh89d02.r
c	2	17.4	60.0	67	9	AA708911	AA708911 zl64a10.s
c	3	16.6	57.2	70	9	AI609394	AI609394 tw93b03.x
c	4	16	55.2	51	10	BG361927	BG361927 gb49d10.y
c	5	15.8	54.5	58	9	AI824019	AI824019 wj29f03.x
	6	15.6	53.8	37	9	AI802260	AI802260 tj36g07.x
c	7	15.6	53.8	58	28	AZ834846	AZ834846 2M0117F18
	8	15.4	53.1	49	14	U44334	U44334 ENU44334 As
	9	15.2	52.4	61	9	AI318033	AI318033 ta75g02.x
	10	15.2	52.4	65	12	BM517546	BM517546 kj80g07.y
c	11	15	51.7	58	28	B02943	B02943 cSRL-163G2-
c	12	14.8	51.0	34	28	AZ840876	AZ840876 2M0138C08
c	13	14.8	51.0	35	9	AL801069	AL801069 AL801069
	14	14.8	51.0	49	28	AZ576537	AZ576537 AST-T11C0
c	15	14.8	51.0	55	9	AI224478	AI224478 qx06d06.x
	16	14.8	51.0	64	10	BE636255	BE636255 SWOvAMCAQ
	17	14.8	51.0	65	9	AL895107	AL895107 AL895107
	18	14.6	50.3	53	29	AL940874	AL940874 Arabidops
	19	14.6	50.3	59	10	BE970792	BE970792 601680150
c	20	14.6	50.3	61	13	BQ479345	BQ479345 ku33d12.y
c	21	14.6	50.3	65	29	AL763793	AL763793 Arabidops
c	22	14.6	50.3	69	29	BZ768797	BZ768797 SALK_1407
c	23	14.6	50.3	70	29	BZ768791	BZ768791 SALK_1407
c	24	14.6	50.3	70	29	BZ768795	BZ768795 SALK_1407
	25	14.4	49.7	35	28	BH856246	BH856246 SALK_0811
	26	14.4	49.7	35	28	BH856247	BH856247 SALK_0811
c	27	14.4	49.7	37	28	AZ950243	AZ950243 2M0214C15
c	28	14.4	49.7	41	28	AZ598587	AZ598587 1M0413A04
	29	14.4	49.7	51	29	DME545740	AJ545740 Drosophil
	30	14.4	49.7	56	29	BZ665747	BZ665747 KG10262 D
c	31	14.4	49.7	57	10	BG362067	BG362067 gb47b08.y
c	32	14.4	49.7	58	9	AV953887	AV953887 AV953887
	33	14.4	49.7	64	9	AI321110	AI321110 d4c09nm.r
c	34	14.4	49.7	65	28	BH908271	BH908271 SALK_0468
c	35	14.4	49.7	66	10	BG361679	BG361679 gb48b04.y
c	36	14.4	49.7	67	28	BH848343	BH848343 SALK_0678
c	37	14.4	49.7	68	10	BG362185	BG362185 gb52h02.y
	38	14.4	49.7	70	28	BH759592	BH759592 KG05236-3
c	39	14.2	49.0	38	29	BZ355014	BZ355014 SALK_1262
c	40	14.2	49.0	54	28	B05408	B05408 cSRL-62a5-u
	41	14.2	49.0	56	28	AZ938752	AZ938752 2M0197N19
c	42	14.2	49.0	64	29	BX161966	BX161966 Danio rer
c	43	14.2	49.0	65	9	AI719509	AI719509 as44b06.x
c	44	14.2	49.0	65	13	BQ564818	BQ564818 gi24h01.y
c	45	14.2	49.0	69	9	AI211081	AI211081 n0e06a1.f

ALIGNMENTS

RESULT 1
AA516989/c
LOCUS AA516989 70 bp mRNA linear EST 14-JUL-1997
DEFINITION vh89d02.r1 Knowles Solter mouse embryonic stem cell Mus musculus cDNA clone IMAGE:894147 5' similar to TR:G187568 G187568 MG44 ;, mRNA sequence.
ACCESSION AA516989
VERSION AA516989.1 GI:2256448
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 70)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine?
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

```
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:522107
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            High quality sequence stop: 1.
FEATURES              Location/Qualifiers
     source           1. .70
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6J x DBA/2J F1"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:894147"
                      /dev_stage="embryo"
                      /lab_host="DH10B"
                      /clone_lib="Knowles Solter mouse embryonic stem cell"
                      /note="Vector: pSPORT; Site_1: NotI; Site_2: SalI; Cloned
                      unidirectionally from mRNA prepared from 800 blastocysts.
                      Primer: SalI(dT): 5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTT-3'.
                      cDNAs  were cloned into the NotI/SalI sites of a pSPORT
                      vector (Life  Technologies).  "
BASE COUNT       16 a     14 c     15 g     25 t
ORIGIN

  Query Match             63.4%;  Score 18.4;  DB 9;  Length 70;
  Best Local Similarity   57.1%;  Pred. No. 1.4e+04;
  Matches   16;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AAAGAUUCUUUUUGUAAGCCCCAAGGGC 28
              | |||::|:: ::  || | ||||||||
Db         47 ACAGATTCTTCTTAGAAACACCAAGGGC 20


RESULT 2
AA708911/c
LOCUS       AA708911                 67 bp    mRNA    linear   EST 24-DEC-1997
DEFINITION  zl64a10.s1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:506682 3' similar to SW:RB32_HUMAN Q13637 RAS-RELATED PROTEIN
            RAB-32 ;, mRNA sequence.
ACCESSION   AA708911
VERSION     AA708911.1  GI:2718829
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 67)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES              Location/Qualifiers
     source           1. .67
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="GDB:3812701"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:506682"
                      /sex="female"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /clone_lib="Soares_pregnant_uterus_NbHPU"
                      /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization.  Library
                      constructed by M. Fatima Bonaldo."
BASE COUNT       16 a     16 c     18 g     17 t
ORIGIN

  Query Match             60.0%;  Score 17.4;  DB 9;  Length 67;
  Best Local Similarity   55.6%;  Pred. No. 3.4e+04;
  Matches   15;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 AGAUUCUUUUUGUAAGCCCCAAGGGCU 29
              |||  |:: ::|:|| |||| | |||:
Db         51 AGAGACTTCTTGTAAACCCCCAAGGCT 25


RESULT 3
AI609394/c
LOCUS       AI609394                 70 bp    mRNA    linear   EST 16-DEC-1999
DEFINITION  tw93b03.x1 NCI_CGAP_HN6 Homo sapiens cDNA clone IMAGE:2267213 3'
            similar to SW:TCPH_HUMAN Q99832 T-COMPLEX PROTEIN 1, ETA SUBUNIT ;,
            mRNA sequence.
ACCESSION   AI609394
VERSION     AI609394.1  GI:4618561
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 70)
  AUTHORS   NCI/NIDR-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Dental Research,
            Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Chong Heon Lee, D.D.S., Mary May, J. Silvio
            Gutkind, Ph.D., Myung Hee Park, PhD.
             cDNA Library Preparation: Stratagene, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 2028   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .70
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2267213"
                      /tissue_type="normal gingiva (cell line from immortalized
                      keratinocytes)"
                      /lab_host="SOLR (kanamycin resistant)"
                      /clone_lib="NCI_CGAP_HN6"
                      /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                      ; Cloned unidirectionally.  Primer: Oligo dT. Average
                      insert size 1.3 kb. 5' adaptor sequence: 5' AATTCGGCACGAG
                      3'                      3'     GCCGTGCTC 5' 3' adaptor
```

sequence: 5' (GA)10ACTAGTCTCGAGTTTTTTTTTTTTTTTT 3' EcoRI site appears to have been lost in a fraction of the clones. Library constructed by Stratagene; available through Mary May, PhD (Oral and Pharyngeal Cancer Branch, National Institute of Dental and Craniofacial Research, NIH; mmay@yoda.nidr.nih.gov)."
BASE COUNT 18 a 23 c 14 g 12 t 3 others
ORIGIN

Query Match 57.2%; Score 16.6; DB 9; Length 70;
Best Local Similarity 47.8%; Pred. No. 6.6e+04;
Matches 11; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

```
Qy          6 UUCUUUUUGUAAGCCCCAAGGGC 28
              :: ::::: :  |||||||||  |
Db         54 TTTTTTTTGTGGGCCCCAAGGCC 32
```

RESULT 4
BG361927/c
LOCUS BG361927 51 bp mRNA linear EST 08-MAR-2001
DEFINITION gb49d10.y1 Moss EST library PPG Physcomitrella patens cDNA clone PEP_SOURCE_ID: 5', mRNA sequence.
ACCESSION BG361927
VERSION BG361927.1 GI:13251024
KEYWORDS EST.
SOURCE Physcomitrella patens
ORGANISM Physcomitrella patens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE 1 (bases 1 to 51)
AUTHORS Quatrano,R., Bashiardes,S., Cove,D., Cuming,A., Knight,C., Clifton ,S., Marra,M., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood ,K., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Steptoe,M., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R. , Waterston,R. and Wilson,R.
TITLE Leeds/Wash U Moss EST Project
JOURNAL Unpublished
COMMENT Contact: Ralph Quatrano
Leeds/Wash U Moss EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Libraries were constructed by Dr. Stavros Bashiardes as part of the Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and Washington Univ. in St. Louis (USA) DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
Seq primer: -40RP from Gibco.
FEATURES Location/Qualifiers
source 1. .51
/organism="Physcomitrella patens"
/mol_type="mRNA"
/db_xref="taxon:3218"
/clone="PEP_SOURCE_ID:"
/tissue_type="gametophore: 30 day old tissue, ammonium-grown"
/lab_host="DH10B"
/clone_lib="Moss EST library PPG"
/note="Vector: pAMP1; Construction of the cDNA library was performed by Dr. W. Gregg Clark using a modification of the cDNA synthesis protocol developed in the laboratory of Dr. Michael Lovett by Dr. Yulia Korshunova (personal communication). First polyA + RNA was isolated from total gametophore RNA using oligo dT magnetic beads. Following this, first strand cDNA synthesis was performed on the bead-bound polyA + RNA, during which an oligonucleotide anchor sequence was incorporated onto the 5'-ends of the cDNA. PCR amplification was then used to synthesize the second strand, to amplify the double stranded DNA, and to incorporate dUTP containing sequences into the ends of the double stranded cDNA. This DNA was size selected and cloned into pAMP1 using the CloneAMP pAMP1 System (Life Technologies, GibcoBRL) for cloning amplification products by a non-restriction site dependant process. The cloning was directional based on sequence asymmetry introduced at the ends during PCR amplification. The 3' cDNA ends are proximal to the NotI site of the multiple cloning site in pAMP1. This annealing mixture was transformed into chemically competent DH10B cells and selected for ampicillin resistant growth. The resulting clones (about 330,000) were pooled to make the library."
BASE COUNT 18 a 9 c 8 g 16 t
ORIGIN

Query Match 55.2%; Score 16; DB 10; Length 51;
Best Local Similarity 41.7%; Pred. No. 1.1e+05;
Matches 10; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

```
Qy          6 UUCUUUUUGUAAGCCCCAAGGGCU 29
              :: ::::: :||| |||||| |:
Db         27 TTTTTTTTTTAAGACCCAAGAACT 4
```

RESULT 5
AI824019/c
LOCUS AI824019 58 bp mRNA linear EST 21-DEC-1999
DEFINITION wj29f03.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2404253 3' similar to TR:O70278 O70278 MULTIPLE ENDOCRINE NEOPLASIA TYPE 1 CANDIDATE PROTEIN NUMBER 18. ;, mRNA sequence.
ACCESSION AI824019
VERSION AI824019.1 GI:5444690
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 58)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert Length: 806 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1. .58
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2404253"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Kid12"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_Kid5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1323912-1325831, 1471368-1472903 and

1492104-1493255). Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 11 a 14 c 19 g 14 t
ORIGIN

Query Match 54.5%; Score 15.8; DB 9; Length 58;
Best Local Similarity 44.4%; Pred. No. 1.3e+05;
Matches 12; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

```
Qy          3 AGAUUCUUUUUGUAAGCCCCAAGGGCU 29
              || :: :::::  ||| || ||| ||:
Db         56 AGCTTTTTTTTCCAAGTCCAAAGAGCT 30
```

RESULT 6
AI802260
LOCUS AI802260 37 bp mRNA linear EST 13-DEC-1999
DEFINITION tj36g07.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2143644 3' similar to TR:Q41120 Q41120 HYDROXYPROLINE-RICH GLYCOPROTEIN ;, mRNA sequence.
ACCESSION AI802260
VERSION AI802260.1 GI:5367732
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 37)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert Length: 1470 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1. .37
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2143644"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Pan1"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"
BASE COUNT 6 a 17 c 3 g 11 t
ORIGIN

Query Match 53.8%; Score 15.6; DB 9; Length 37;
Best Local Similarity 50.0%; Pred. No. 1.6e+05;
Matches 11; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 AAAGAUUCUUUUUGUAAGCCCC 22
              |||  :: :::::| |||||||
Db          7 AAAATTTTTTTTTGGAAGCCCC 28
```

RESULT 7
AZ834846/c
LOCUS AZ834846 58 bp DNA linear GSS 20-FEB-2001
DEFINITION 2M0117F18R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0117F18 R, genomic survey sequence.
ACCESSION AZ834846
VERSION AZ834846.1 GI:13004754
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 58)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0117 row: F column: 18
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 58.
FEATURES Location/Qualifiers
source 1. .58
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0117F18"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 15 a 13 c 16 g 14 t
ORIGIN

Query Match 53.8%; Score 15.6; DB 28; Length 58;
Best Local Similarity 54.5%; Pred. No. 1.5e+05;
Matches 12; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

```
Qy          4 GAUUCUUUUUGUAAGCCCCAAG 25
              | ::|  :::|:|| |||||||
Db         24 GTTTCCCTTTGTAATCCCCAAG 3
```

RESULT 8
U44334
LOCUS U44334 49 bp mRNA linear EST 03-APR-1996

```
DEFINITION  ENU44334 Aspergillus nidulans cleistothecium Emericella nidulans
            cDNA clone SE0762, mRNA sequence.
ACCESSION   U44334
VERSION     U44334.1  GI:1244997
KEYWORDS    EST.
SOURCE      Emericella nidulans (anamorph: Aspergillus nidulans)
  ORGANISM  Emericella nidulans
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; Emericella.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   Lee,D., Lee,S., Hwang,H., Kim,J. and Chae,K.
  TITLE     Quantitative analysis of gene expression in sexual structures of
            Aspergillus nidulans by sequencing of 3'-directed cDNA clones
  JOURNAL   FEMS Microbiol. Lett. 138 (1), 71-76 (1996)
  MEDLINE   96236220
   PUBMED   8674973
COMMENT     Contact: Keon-Sang Chae
            Chonbuk National University
            Chonju, 561-756, S. Korea
            Tel: +82-652-70-3340
            Fax: +82-652-70-3345
            Email: chaeks@chonbuknms.chonbuk.ac.kr.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Emericella nidulans"
                     /mol_type="mRNA"
                     /strain="FGSC4"
                     /db_xref="taxon:162425"
                     /clone="SE0762"
                     /tissue_type="cleistothecium"
                     /cell_type="Hull cell"
                     /dev_stage="sexual"
                     /clone_lib="Aspergillus nidulans cleistothecium"
                     /note="3'-directed cDNA clones; single-pass sequencing"
BASE COUNT       11 a     13 c     10 g     15 t
ORIGIN

  Query Match             53.1%;  Score 15.4;  DB 14;  Length 49;
  Best Local Similarity   52.0%;  Pred. No. 1.9e+05;
  Matches   13;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 AGAUUCUUUUUGUAAGCCCCAAGGG 27
              |||::|:::   :||  ||||| ||
Db         20 AGATTCTTTCATTAACTCCCAACGG 44


RESULT 9
AI318033
LOCUS       AI318033                 61 bp    mRNA    linear   EST 16-FEB-1999
DEFINITION  ta75g02.x1 NCI_CGAP_HSC2 Homo sapiens cDNA clone IMAGE:2049938 3'
            similar to SW:RL34_HUMAN P49207 60S RIBOSOMAL PROTEIN L34. ;, mRNA
            sequence.
ACCESSION   AI318033
VERSION     AI318033.1  GI:4033793
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 61)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Herbert Morse, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 384    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .61
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2049938"
                     /tissue_type="stem cell 34+/38+"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_HSC2"
                     /note="Organ: bone marrow; Vector: pAMP1; mRNA made from
                     bone marrow, stem cells 34+/38+, cDNA made by oligo-dT
                     priming.  Directionally cloned.  Size-selected on agarose
                     gel, average insert size 400 bp.  Primary library,
                     non-amplified.  "
BASE COUNT       20 a     10 c     16 g     15 t
ORIGIN

  Query Match             52.4%;  Score 15.2;  DB 9;  Length 61;
  Best Local Similarity   53.6%;  Pred. No. 2.2e+05;
  Matches   15;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 AAAGAUUCUUUUUGUAAGCCCCAAGGGC 28
              || | ::| :  :|:| | || ||||||
Db          1 AAGGGTTCGTGCTGTATGACCTAAGGGC 28


RESULT 10
BM517546
LOCUS       BM517546                 65 bp    mRNA    linear   EST 15-FEB-2002
DEFINITION  kj80g07.y1 Ascaris suum female head SL1 TOPO v1 Murphy Chiapelli
            McCarter Ascaris suum cDNA 5', mRNA sequence.
ACCESSION   BM517546
VERSION     BM517546.1  GI:18688698
KEYWORDS    EST.
SOURCE      Ascaris suum (pig roundworm)
  ORGANISM  Ascaris suum
            Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea
            ; Ascarididae; Ascaris.
REFERENCE   1  (bases 1 to 65)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.
            , Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.
            , Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R.
            , Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
            ,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.
            , Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
            Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Claire Murphy, Brandi Chiapelli, and
            Dr. James McCarter at Washington University, St. Louis. DNA
            Sequencing by: Washington University Genome Sequencing Center.
FEATURES             Location/Qualifiers
     source          1. .65
                     /organism="Ascaris suum"
                     /mol_type="mRNA"
                     /db_xref="taxon:6253"
                     /sex="Female"
                     /tissue_type="Head"
                     /dev_stage="Adult"
```

```
                     /lab_host="DH10B"
                     /clone_lib="Ascaris suum female head SL1 TOPO v1 Murphy
                     Chiapelli McCarter"
                     /note="Vector: pCRII-TOPO (Invitrogen); Site_1: EcoRI;
                     Site_2: EcoRI; The library was constructed by Claire
                     Murphy, Brandi Chiapelli,  and Dr. James McCarter at
                     Washington University, St. Louis. Oligo(dT)-SL1  PCR based
                     library. Ascaris suum female head cDNA PCR products of
                     size  >400 nucleotides containing SL1 on the 5' end and
                     oligo(dT) on the 3' end  were non-directionally cloned
                     into pCRII-TOPO(Invitrogen) following the  Topo TA cloning
                     protocol. Dissected nematode tissues were  provided by Dr.
                     Alan Scott (ascott@jhsph.edu) of the School of Public
                     Hygiene  and Public Health at John Hopkins University in
                     Baltimore, MD"
BASE COUNT       24 a     11 c     15 g     15 t
ORIGIN

  Query Match             52.4%;  Score 15.2;  DB 12;  Length 65;
  Best Local Similarity   42.9%;  Pred. No. 2.1e+05;
  Matches   12;  Conservative    8;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          2 AAGAUUCUUUUUGUAAGCCCCAAGGGCU 29
              ||| ::|: ::: : || || |||  |:
Db          1 AAGGTTCTGTTTATGAGACCGAAGATCT 28


RESULT 11
B02943/c
LOCUS       B02943                   58 bp    DNA     linear   GSS 13-JUL-1996
DEFINITION  cSRL-163G2-u cSRL flow sorted Chromosome 11 specific cosmid Homo
            sapiens genomic clone cSRL-163G2, genomic survey sequence.
ACCESSION   B02943
VERSION     B02943.1  GI:1412221
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 58)
  AUTHORS   Evans,G.A., Burbee,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M.,
            Jones,D., Ward,T., Gillilan,E., Schagemann,J., Probst,S. ., Harris
            ,J., DeFord,J., McFarland,J., Burzinski,K., Khan,M., Kupfer,K. and
            Garner,H.R.
  TITLE     Genomic Sequence Sampled Map of Chromosome 11
  JOURNAL   Unpublished
COMMENT     Contact: Evans GA, Shane Probst
            McDermott Center for Human Growth and Development
            University of Texas Southwestern Medical Center At Dallas
            5323 Harry Hines Blvd, Dallas TX 75235-8591
            Tel: 214-648-1600
            Fax: 214-648-1666
            Email: gevans@utsw.swmed.edu, shane@mcdermott.swmed.edu
            Seq primer: T7
            Class: cosmid ends
            High quality sequence stop: 58.
FEATURES             Location/Qualifiers
     source          1. .58
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="cSRL-163G2"
                     /sex="female"
                     /cell_type="chimeric hamster somatic cell hybrid"
                     /clone_lib="cSRL flow sorted Chromosome 11 specific
                     cosmid"
                     /note="Vector: sCos-1; Human Chromosome 11 specific cosmid
                     library prepared from flow sorted human  Chromosome 11
                     derived from Chinese Hampster Ovary (CHO) monochromosomal
                     somatic cell hybrid, J1"
BASE COUNT       25 a      9 c      8 g     15 t      1 others
ORIGIN

  Query Match             51.7%;  Score 15;  DB 28;  Length 58;
  Best Local Similarity   43.5%;  Pred. No. 2.6e+05;
  Matches   10;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          6 UUCUUUUUGUAAGCCCCAAGGGC 28
              :: ::::: : | ||| ||||||
Db         41 TTTTTTTTTTGATCCCAAAGGGC 19


RESULT 12
AZ840876/c
LOCUS       AZ840876                 34 bp    DNA     linear   GSS 20-FEB-2001
DEFINITION  2M0138C08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0138C08 R, genomic survey sequence.
ACCESSION   AZ840876
VERSION     AZ840876.1  GI:13010784
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 34)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0138  row: C  column: 08
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 34.
FEATURES             Location/Qualifiers
     source          1. .34
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0138C08"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        6 a      9 c      8 g     11 t
ORIGIN
```

```
  Query Match             51.0%;  Score 14.8;  DB 28;  Length 34;
  Best Local Similarity   57.7%;  Pred. No. 3.2e+05;
  Matches   15;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          3 AGAUUCUUUUUGUAAGCCCCAAGGGC 28
              | |:  : :: | |||| ||||||||
Db         26 ATATAATCTTCGAAAGCACCAAGGGC 1


RESULT 13
AL801069/c
LOCUS       AL801069                 35 bp    mRNA    linear   EST 27-JUN-2002
DEFINITION  AL801069 XGC-neurula Silurana tropicalis cDNA clone TNeu127i24 5',
            mRNA sequence.
ACCESSION   AL801069
VERSION     AL801069.1  GI:21587437
KEYWORDS    EST.
SOURCE      Silurana tropicalis (western clawed frog)
  ORGANISM  Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE   1  (bases 1 to 35)
  AUTHORS   Taylor,R., Ashurst,J.L., Croning,M.D.R., Zorn,A.M. and Rogers,J.
  TITLE     Sanger Xenopus tropicalis EST project 2002
  JOURNAL   Unpublished
COMMENT     Contact: Taylor R
            Sanger Centre
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: trop@sanger.ac.uk
            Sanger Xenopus tropicalis EST project 2001
            TROPICALIS_SEQUENCE_ID: TNeu127i24.p1cSP6
            Sequencing primer: P1CSP6
            This sequence is from a Xenopus Gene Collection (XGC) library
            constructed by Aaron M. Zorn.
FEATURES             Location/Qualifiers
     source          1. .35
                     /organism="Silurana tropicalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8364"
                     /clone="TNeu127i24"
                     /dev_stage="neurula"
                     /lab_host="Escherichia coli DH10B"
                     /clone_lib="XGC-neurula"
                     /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
                     was oligo dT primed from 5ug of poly A+ RNA from neurula.
                     EcoRI-NotI cut cDNA was then ligated into pCS107 with
                     EcoRI at the 5' end and NotI at the 3' end."
BASE COUNT      13 a      7 c      8 g      7 t
ORIGIN

  Query Match             51.0%;  Score 14.8;  DB 9;  Length 35;
  Best Local Similarity   42.3%;  Pred. No. 3.2e+05;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AAAGAUUCUUUUUGUAAGCCCCAAGG 26
              ||| |:: ::::: :  |||||  ||
Db         31 AAAAATTTTTTTTTTGTGCCCCGGGG 6


RESULT 14
AZ576537
LOCUS       AZ576537                 49 bp    DNA     linear   GSS 06-DEC-2000
DEFINITION  AST-T11C0260\{ Genetrap T47D Human Breast Carcinoma Library Homo
            sapiens genomic 5', genomic survey sequence.
ACCESSION   AZ576537
VERSION     AZ576537.1  GI:11562848
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   Henkel,G., Liyanage,M., Pratt,E., Huang,D., Riley,M., Bernardino,A.
            , Durick,K. and Pollok,B.
  TITLE     Exon-trap tags from a T47D GenomeScreen(TM) Library
  JOURNAL   Unpublished
COMMENT     Contact: Greg Henkel
            Gene Expression
            Aurora Biosciences Corp.
            11010 Torreyana Road, San Diego, CA 92121, USA
            Tel: 8584048436
            Fax: 8584046719
            Email: henkelg@aurorabio.com
            Pools of cells were isolated from a GenomeScreen(TM) library. The
            library of cells was generated by retroviral integration of a gene
            tagging element consisting of: 1) A promoterless beta-lactamase
            proceeded by a splice acceptor as a reporter for gene expression;
            2) A promoter driving neomycin resistance followed by a splice
            donor to trap downstream exons. 3' RACE from neomycin gene was
            performed using total RNA from isolated pools. Output was shotgun
            cloned in pAmp-1 and used to transform DH5-alpha competent
            bacteria. 5' ends of reported sequences were immediately preceded
            by splice donor from the trapping co nstruct.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /tissue_type="Carcinoma"
                     /cell_type="Epithelial"
                     /cell_line="T47D"
                     /clone_lib="Genetrap T47D Human Breast Carcinoma Library"
                     /note="Organ: Breast; Vector: pAmp-1; 3' RACE of total RNA
                     from genetrap pools; shotgun clone in pAmp-1 and  used to
                     transform DH5-alpha competent bacteria."
BASE COUNT      18 a     11 c     14 g      6 t
ORIGIN

  Query Match             51.0%;  Score 14.8;  DB 28;  Length 49;
  Best Local Similarity   57.7%;  Pred. No. 3.1e+05;
  Matches   15;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AAAGAUUCUUUUUGUAAGCCCCAAGG 26
              |||||  |::  :|:||||  ||| |
Db          6 AAAGAGACTTCCTGTAAGCGGCAAAG 31


RESULT 15
AI224478/c
LOCUS       AI224478                 55 bp    mRNA    linear   EST 21-DEC-1998
DEFINITION  qx06d06.x1 NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2000555 3'
            similar to TR:O23462 O23462 HYPOTHETICAL 18.0 KD PROTEIN. ;, mRNA
            sequence.
ACCESSION   AI224478
VERSION     AI224478.1  GI:3807191
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 55)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            unknown library type
            Trace considered overall poor quality
            Insert Length: 1214   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
```

```
     source          1. .55
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2000555"
                     /tissue_type="lymphoma, follicular mixed small and large
                     cell"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Lym12"
                     /note="Organ: lymph node; Vector: pCMV-SPORT6; Site_1:
                     SalI; Site_2: NotI; Cloned unidirectionally.  Primer:
                     Oligo dT.  Average insert size 1.25 kb. Life Technologies
                     catalog #: 11547-015"
BASE COUNT       19 a     18 c     14 g      4 t
ORIGIN

  Query Match             51.0%;  Score 14.8;  DB 9;  Length 55;
  Best Local Similarity   38.5%;  Pred. No. 3e+05;
  Matches   10;  Conservative    9;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCCAAGGGCU 29
              | :: ::::: :   ||||||||| :
Db         30 GTTTTTTTTTTTTCCCCCAAGGGTT 5


Search completed: January 30, 2004, 10:12:41
Job time : 1621 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 07:56:58 ; Search time 50 Seconds
(without alignments)
256.002 Million cell updates/sec

Title: US-09-310-844C-24
Perfect score: 29
Sequence: 1 uaugauucuuuuuguaagcccuaggggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 792150

Minimum DB seq length: 0
Maximum DB seq length: 70

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	15.2	52.4	25	3	US-08-943-731-336	Sequence 336, App
c	2	15.2	52.4	33	1	US-08-667-079B-5	Sequence 5, Appli
c	3	14.8	51.0	35	6	5422260-12	Patent No. 5422260
c	4	14.8	51.0	36	3	US-09-440-001-1	Sequence 1, Appli
c	5	14.8	51.0	36	4	US-09-605-685-1	Sequence 1, Appli
	6	14	48.3	30	4	US-09-690-146A-5	Sequence 5, Appli
c	7	14	48.3	30	4	US-09-690-146A-7	Sequence 7, Appli
	8	14	48.3	37	1	US-08-049-264C-55	Sequence 55, Appl
	9	14	48.3	37	1	US-08-476-562-55	Sequence 55, Appl
	10	14	48.3	37	1	US-08-479-723A-55	Sequence 55, Appl
	11	14	48.3	37	5	PCT-US94-04310-55	Sequence 55, Appl
c	12	14	48.3	44	1	US-08-049-264C-54	Sequence 54, Appl
c	13	14	48.3	44	1	US-08-476-562-54	Sequence 54, Appl
c	14	14	48.3	44	1	US-08-479-723A-54	Sequence 54, Appl
c	15	14	48.3	44	5	PCT-US94-04310-54	Sequence 54, Appl
	16	14	48.3	47	4	US-09-641-638-1059	Sequence 1059, Ap
c	17	13.8	47.6	36	3	US-09-440-001-3	Sequence 3, Appli
c	18	13.8	47.6	36	4	US-09-605-685-3	Sequence 3, Appli
	19	13.8	47.6	47	4	US-09-422-978-639	Sequence 639, App
c	20	13.6	46.9	42	4	US-09-468-872-11	Sequence 11, Appl
	21	13.6	46.9	44	2	US-08-343-443B-39	Sequence 39, Appl
c	22	13.6	46.9	47	4	US-09-422-978-2286	Sequence 2286, Ap
	23	13.4	46.2	26	3	US-09-247-190-37	Sequence 37, Appl
c	24	13.4	46.2	35	4	US-09-598-747-32	Sequence 32, Appl
	25	13.4	46.2	41	1	US-08-468-220-28	Sequence 28, Appl
	26	13.4	46.2	41	2	US-08-468-698-28	Sequence 28, Appl
	27	13.4	46.2	41	3	US-08-194-664A-28	Sequence 28, Appl
	28	13.4	46.2	41	5	PCT-US94-01553A-28	Sequence 28, Appl
	29	13.4	46.2	41	5	PCT-US95-10426-28	Sequence 28, Appl
c	30	13.4	46.2	51	1	US-08-328-152A-11	Sequence 11, Appl
	31	13.4	46.2	52	4	US-09-310-463-6	Sequence 6, Appli
	32	13.4	46.2	52	4	US-08-842-248A-6	Sequence 6, Appli
c	33	13.4	46.2	60	3	US-08-478-097A-32	Sequence 32, Appl
c	34	13.4	46.2	60	4	US-09-496-398-32	Sequence 32, Appl
c	35	13.2	45.5	19	3	US-08-532-896-53	Sequence 53, Appl
	36	13.2	45.5	27	3	US-09-106-182-17	Sequence 17, Appl
	37	13.2	45.5	27	3	US-09-106-182-23	Sequence 23, Appl
	38	13.2	45.5	27	4	US-09-227-357-4	Sequence 4, Appli
	39	13.2	45.5	27	4	US-09-280-839-7	Sequence 7, Appli
	40	13.2	45.5	27	4	US-09-411-977-19	Sequence 19, Appl
	41	13.2	45.5	27	4	US-09-479-729B-24	Sequence 24, Appl
	42	13.2	45.5	27	4	US-09-257-179-4	Sequence 4, Appli
	43	13.2	45.5	27	4	US-09-149-476-4	Sequence 4, Appli
	44	13.2	45.5	27	4	US-09-288-143-4	Sequence 4, Appli
	45	13.2	45.5	27	4	US-09-487-792-26	Sequence 26, Appl

ALIGNMENTS

```
RESULT 1
US-08-943-731-336/c
; Sequence 336, Application US/08943731
; Patent No. 6265157
; GENERAL INFORMATION:
;   APPLICANT: PROCKOP, DARWIN J.
;   APPLICANT: SPOTILA, LORETTA D.
;   APPLICANT: DELTAS, CONSTANTINOS D.
;   APPLICANT: SEREDA, LARISA
;   APPLICANT: LARSON, ANDREA W.
;   APPLICANT: PACK, MICHAEL
;   APPLICANT: COLIGE, ALAIN
;   APPLICANT: EARLY, JAMES
;   APPLICANT: KORKKO, JARMO
;   APPLICANT: ALA-KOKKO, LEENA, et al.
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
;   TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
;   NUMBER OF SEQUENCES: 666
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
;     STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
;     STREET: FLR.
;     CITY: PHILADELPHIA
;     STATE: PA
;     COUNTRY: USA
;     ZIP: 19103-7086
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/943,731
;     FILING DATE: 03-OCT-1997
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/212,322
;     FILING DATE: 14-MAR-1994
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/803,628
;     FILING DATE: 03-DEC-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME: DOYLE LEARY Ph.D., KATHRYN
;     REGISTRATION NUMBER: 36,317
;     REFERENCE/DOCKET NUMBER: 9598-27
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 215-965-1284
;     TELEFAX: 215-567-2991
;     TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 336:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  25 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-943-731-336

  Query Match             52.4%;  Score 15.2;  DB 3;  Length 25;
  Best Local Similarity   45.0%;  Pred. No. 2.7e+02;
  Matches    9;  Conservative    8;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          5 AUUCUUUUUGUAAGCCCUAG 24
              |::|: :::|: |||||: |
Db         22 ATTCTCTTTGTGAGCCCTTG 3


RESULT 2
US-08-667-079B-5/c
; Sequence 5, Application US/08667079B
; Patent No. 5789171
;  GENERAL INFORMATION:
;    APPLICANT:  Mark S. Smeltzer
;    TITLE OF INVENTION:  Use of cna, fnbA, fnbB, and hlb Gene Probes for the Strain-Sp
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Benjamin Aaron Adler, MCGREGOR & ADLER, P.C.
;      STREET:  8011 Candle Lane
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word for Macintosh
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/667,079B
;      FILING DATE:  June 20, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adler, Benjamin Aaron
;      REGISTRATION NUMBER:  35,423
;      REFERENCE/DOCKET NUMBER:  D5886
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-777-2321
;      TELEFAX:  713-777-6908
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  other nucleic acid
;    HYPOTHETICAL:  No
;    ANTI-SENSE:  No
;    ORIGINAL SOURCE:
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:
US-08-667-079B-5

  Query Match             52.4%;  Score 15.2;  DB 1;  Length 33;
  Best Local Similarity   39.3%;  Pred. No. 2.8e+02;
  Matches   11;  Conservative    9;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          2 AUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
              |:||:: :::: |:||   |   ||||:
Db         32 ATGATTGTTTTAGTAATTTCCCCGGGCT 5


RESULT 3
5422260-12/c
;Patent No. 5422260
;    APPLICANT: KAUFMAN, RANDAL J.;PITTMAN, DEBRA D.;TOOLE,JOHN J.
;    TITLE OF INVENTION: HUMAN FACTOR VIII: C MUTEINS
;    NUMBER OF SEQUENCES: 15
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/883,936
;      FILING DATE: 15-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 279,485
;      FILING DATE: 02-DEC-1988; 09-DEC-1986
;      APPLICATION NUMBER: 939,658
;      FILING DATE: 09-DEC-1986
;      APPLICATION NUMBER: 932,767
;      FILING DATE: 18-NOV-1986
;      APPLICATION NUMBER: 868,410
;      FILING DATE: 29-MAY-1986
;SEQ ID NO:12:
;      LENGTH: 35
5422260-12

  Query Match             51.0%;  Score 14.8;  DB 6;  Length 35;
  Best Local Similarity   42.3%;  Pred. No. 4.4e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCUAGGGGCU 29
              | ::  ::::| |||| : |||||:
Db         35 GTTTTCTTTTGAAAGCTTTTGGGGCT 10


RESULT 4
US-09-440-001-1/c
; Sequence 1, Application US/09440001
; Patent No. 6174696
; GENERAL INFORMATION:
;  APPLICANT:  Seman, Leo J.
;  TITLE OF INVENTION:  A METHOD FOR THE DETERMINATION OF HOMOCYSTEINE
;  FILE REFERENCE:  09/440,001
;  CURRENT APPLICATION NUMBER: US/09/440,001
;  CURRENT FILING DATE:  1999-11-12
;  PRIOR APPLICATION NUMBER:  60/108,099
;  PRIOR FILING DATE:  1998-11-12
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 36
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: Oligonucleotide primer
US-09-440-001-1

  Query Match             51.0%;  Score 14.8;  DB 3;  Length 36;
  Best Local Similarity   38.5%;  Pred. No. 4.4e+02;
  Matches   10;  Conservative    9;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGG 26
              :|: |  |:::::|:  ||| :| ||
Db         33 TATCAAGCTTTTTGTCCGCCATATGG 8


RESULT 5
US-09-605-685-1/c
; Sequence 1, Application US/09605685
; Patent No. 6436658
; GENERAL INFORMATION:
;  APPLICANT: Seman, Leo J.
;  TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF HOMOCYSTEINE
```

```
;  FILE REFERENCE: 09/440,001
;  CURRENT APPLICATION NUMBER: US/09/605,685
;  CURRENT FILING DATE:  2000-06-26
;  PRIOR APPLICATION NUMBER: 60/108,099
;  PRIOR FILING DATE: 1998-11-12
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 36
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: Oligonucleotide primer
US-09-605-685-1

  Query Match             51.0%;  Score 14.8;  DB 4;  Length 36;
  Best Local Similarity   38.5%;  Pred. No. 4.4e+02;
  Matches   10;  Conservative    9;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGG 26
              :|: |  |:::::|:  ||| :| ||
Db         33 TATCAAGCTTTTTGTCCGCCATATGG 8


RESULT 6
US-09-690-146A-5
; Sequence 5, Application US/09690146A
; Patent No. 6485937
; GENERAL INFORMATION:
;  APPLICANT: Palhan, Vikas
;  APPLICANT:  Roeder, Robert
;  TITLE OF INVENTION: System for Rapid Generation of Recombinant
;  TITLE OF INVENTION:  Baculovirus-Based Expression Vectors for Silkworm Larvae
;  FILE REFERENCE: 7529/1G164-US1
;  CURRENT APPLICATION NUMBER: US/09/690,146A
;  CURRENT FILING DATE:  2001-06-01
;  PRIOR APPLICATION NUMBER: 60/159,707
;  PRIOR FILING DATE: 1999-10-15
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
;   LENGTH: 30
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: VP28 Reverse Primer
US-09-690-146A-5

  Query Match             48.3%;  Score 14;  DB 4;  Length 30;
  Best Local Similarity   45.5%;  Pred. No. 1e+03;
  Matches   10;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          5 AUUCUUUUUGUAAGCCCUAGGG 26
              |::   :::|:|| || :|||
Db          2 ATTAAATTTGTAATCCTTAGGG 23


RESULT 7
US-09-690-146A-7/c
; Sequence 7, Application US/09690146A
; Patent No. 6485937
; GENERAL INFORMATION:
;  APPLICANT: Palhan, Vikas
;  APPLICANT:  Roeder, Robert
;  TITLE OF INVENTION: System for Rapid Generation of Recombinant
;  TITLE OF INVENTION:  Baculovirus-Based Expression Vectors for Silkworm Larvae
;  FILE REFERENCE: 7529/1G164-US1
;  CURRENT APPLICATION NUMBER: US/09/690,146A
;  CURRENT FILING DATE:  2001-06-01
;  PRIOR APPLICATION NUMBER: 60/159,707
;  PRIOR FILING DATE: 1999-10-15
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
;   LENGTH: 30
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthesized oligonucleotide
US-09-690-146A-7

  Query Match             48.3%;  Score 14;  DB 4;  Length 30;
  Best Local Similarity   45.5%;  Pred. No. 1e+03;
  Matches   10;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          5 AUUCUUUUUGUAAGCCCUAGGG 26
              |::   :::|:|| || :|||
Db         29 ATTAAATTTGTAATCCTTAGGG 8


RESULT 8
US-08-049-264C-55
; Sequence 55, Application US/08049264C
; Patent No. 5518901
;  GENERAL INFORMATION:
;    APPLICANT:  Murtagh, James J.
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg., 127
;      STREET:  Peachtree Street N.E.
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/049,264C
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perryman, David G.
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  1313.001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 688-0770
;      TELEFAX:  (404) 688-9880
;  INFORMATION FOR SEQ ID NO:  55:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-049-264C-55

  Query Match             48.3%;  Score 14;  DB 1;  Length 37;
  Best Local Similarity   40.9%;  Pred. No. 1e+03;
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          6 UUCUUUUUGUAAGCCCUAGGGG 27
              :: ::::: :|| |||  ||||
Db          8 TTTTTTTTTTAAACCCGGGGGG 29


RESULT 9
US-08-476-562-55
; Sequence 55, Application US/08476562
; Patent No. 5688669
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Murtagh, James J.
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg., 127
;      STREET:  Peachtree Street N.E.
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/476,562
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/049,264
;      FILING DATE:  April 19, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perryman, David G.
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  1313.004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 688-0770
;      TELEFAX:  (404) 688-9880
;  INFORMATION FOR SEQ ID NO:  55:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-476-562-55

  Query Match             48.3%;  Score 14;  DB 1;  Length 37;
  Best Local Similarity   40.9%;  Pred. No. 1e+03;
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          6 UUCUUUUUGUAAGCCCUAGGGG 27
              :: ::::: :|| |||  ||||
Db          8 TTTTTTTTTTAAACCCGGGGGG 29


RESULT 10
US-08-479-723A-55
; Sequence 55, Application US/08479723A
; Patent No. 5744306
;  GENERAL INFORMATION:
;    APPLICANT:  Murtagh, James J.
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg., 127
;      STREET:  Peachtree Street N.E.
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,723A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perryman, David G.
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  05010.0061
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 688-0770
;      TELEFAX:  (404) 688-9880
;  INFORMATION FOR SEQ ID NO:  55:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  oligonucleotide
US-08-479-723A-55

  Query Match             48.3%;  Score 14;  DB 1;  Length 37;
  Best Local Similarity   40.9%;  Pred. No. 1e+03;
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          6 UUCUUUUUGUAAGCCCUAGGGG 27
              :: ::::: :|| |||  ||||
Db          8 TTTTTTTTTTAAACCCGGGGGG 29


RESULT 11
PCT-US94-04310-55
; Sequence 55, Application PC/TUS9404310
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  74
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/04310
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/049,264
;      FILING DATE:  19-APR-1993
;  INFORMATION FOR SEQ ID NO:  55:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US94-04310-55

  Query Match             48.3%;  Score 14;  DB 5;  Length 37;
  Best Local Similarity   40.9%;  Pred. No. 1e+03;
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          6 UUCUUUUUGUAAGCCCUAGGGG 27
              :: ::::: :|| |||  ||||
Db          8 TTTTTTTTTTAAACCCGGGGGG 29


RESULT 12
US-08-049-264C-54/c
; Sequence 54, Application US/08049264C
; Patent No. 5518901
;  GENERAL INFORMATION:
;    APPLICANT:  Murtagh, James J.
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  75
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg., 127
;      STREET:  Peachtree Street N.E.
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/049,264C
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perryman, David G.
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  1313.001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 688-0770
;      TELEFAX:  (404) 688-9880
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-049-264C-54

  Query Match             48.3%;  Score 14;  DB 1;  Length 44;
  Best Local Similarity   40.9%;  Pred. No. 1.1e+03;
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          6 UUCUUUUUGUAAGCCCUAGGGG 27
              :: ::::: :|| |||  ||||
Db         42 TTTTTTTTTTAAACCCGGGGGG 21


RESULT 13
US-08-476-562-54/c
; Sequence 54, Application US/08476562
; Patent No. 5688669
;  GENERAL INFORMATION:
;    APPLICANT:  Murtagh, James J.
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg., 127
;      STREET:  Peachtree Street N.E.
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/476,562
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/049,264
;      FILING DATE:  April 19, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perryman, David G.
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  1313.004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 688-0770
;      TELEFAX:  (404) 688-9880
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-476-562-54

  Query Match             48.3%;  Score 14;  DB 1;  Length 44;
  Best Local Similarity   40.9%;  Pred. No. 1.1e+03;
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          6 UUCUUUUUGUAAGCCCUAGGGG 27
              :: ::::: :|| |||  ||||
Db         42 TTTTTTTTTTAAACCCGGGGGG 21


RESULT 14
US-08-479-723A-54/c
; Sequence 54, Application US/08479723A
; Patent No. 5744306
;  GENERAL INFORMATION:
;    APPLICANT:  Murtagh, James J.
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg., 127
;      STREET:  Peachtree Street N.E.
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,723A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perryman, David G.
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  05010.0061
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 688-0770
;      TELEFAX:  (404) 688-9880
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  oligonucleotide
US-08-479-723A-54

  Query Match             48.3%;  Score 14;  DB 1;  Length 44;
  Best Local Similarity   40.9%;  Pred. No. 1.1e+03;
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          6 UUCUUUUUGUAAGCCCUAGGGG 27
              :: ::::: :|| |||  ||||
Db         42 TTTTTTTTTTAAACCCGGGGGG 21
```

```
RESULT 15
PCT-US94-04310-54/c
; Sequence 54, Application PC/TUS9404310
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  74
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/04310
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/049,264
;      FILING DATE:  19-APR-1993
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US94-04310-54

  Query Match             48.3%;  Score 14;  DB 5;  Length 44;
  Best Local Similarity   40.9%;  Pred. No. 1.1e+03;
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          6 UUCUUUUUGUAAGCCCUAGGGG 27
              :: ::::: :|| |||  ||||
Db         42 TTTTTTTTTTAAACCCGGGGGG 21


Search completed: January 30, 2004, 10:15:13
Job time : 53 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 06:19:17 ; Search time 283.333 Seconds
(without alignments)
276.295 Million cell updates/sec

Title: US-09-310-844C-24
Perfect score: 29
Sequence: 1 uaugauucuuuuuguaagcccuaggggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 2640686

Minimum DB seq length: 0
Maximum DB seq length: 70

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	29	100.0	29	21	AAA70828	Molecular interact
2	29	100.0	42	21	AAA71123	Molecular interact
3	29	100.0	42	21	AAA71131	Molecular interact
4	28	96.6	45	21	AAA70824	Molecular interact
5	28	96.6	46	21	AAA71087	Molecular interact
6	28	96.6	46	21	AAA71096	Molecular interact
7	28	96.6	46	21	AAA71099	Molecular interact
8	28	96.6	46	21	AAA71100	Molecular interact
9	28	96.6	46	21	AAA71104	Molecular interact
10	25.8	89.0	42	21	AAA71113	Molecular interact
11	25.8	89.0	42	21	AAA71118	Molecular interact
12	25.8	89.0	42	21	AAA71126	Molecular interact
13	24.8	85.5	46	21	AAA71085	Molecular interact
14	24.8	85.5	46	21	AAA71103	Molecular interact
15	23.8	82.1	42	21	AAA71114	Molecular interact
16	23.8	82.1	42	21	AAA71119	Molecular interact
17	23.8	82.1	42	21	AAA71127	Molecular interact
18	23.8	82.1	46	21	AAA71094	Molecular interact
19	23.8	82.1	46	21	AAA71110	Molecular interact
20	23.2	80.0	29	21	AAA70829	Molecular interact
21	23.2	80.0	29	21	AAA70830	Molecular interact
22	23.2	80.0	42	21	AAA71115	Molecular interact
23	23.2	80.0	42	21	AAA71116	Molecular interact
24	23.2	80.0	42	21	AAA71120	Molecular interact
25	23.2	80.0	42	21	AAA71121	Molecular interact
26	23.2	80.0	42	21	AAA71128	Molecular interact
27	23.2	80.0	42	21	AAA71129	Molecular interact
28	22.6	77.9	42	21	AAA71124	Molecular interact
29	22.6	77.9	42	21	AAA71132	Molecular interact
30	22.2	76.6	45	21	AAA70825	Molecular interact
31	22.2	76.6	45	21	AAA70826	Molecular interact
32	22.2	76.6	46	21	AAA71088	Molecular interact
33	22.2	76.6	46	21	AAA71089	Molecular interact
34	22.2	76.6	46	21	AAA71090	Molecular interact
35	22.2	76.6	46	21	AAA71105	Molecular interact
36	22.2	76.6	46	21	AAA71106	Molecular interact
37	22.2	76.6	46	21	AAA71107	Molecular interact
38	21.6	74.5	46	21	AAA71093	Molecular interact
39	21.6	74.5	46	21	AAA71095	Molecular interact
40	21.6	74.5	46	21	AAA71109	Molecular interact
41	21.6	74.5	46	21	AAA71111	Molecular interact
42	19.4	66.9	46	21	AAA71084	Molecular interact
43	19.4	66.9	46	21	AAA71098	Molecular interact
44	19.4	66.9	46	21	AAA71102	Molecular interact
45	18.4	63.4	42	21	AAA71117	Molecular interact

ALIGNMENTS

RESULT 1
AAA70828
ID AAA70828 standard; RNA; 29 BP.
XX
AC AAA70828;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #28.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Homo sapiens.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,

PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Claim 235; Page 235; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 29 BP; 5 A; 5 C; 7 G; 12 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
||||||||||||||||||||||||||||||
Db 1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29

RESULT 2
AAA71123
ID AAA71123 standard; DNA; 42 BP.
XX
AC AAA71123;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #129.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 125; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 9 A; 6 C; 9 G; 18 T; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 58.6%; Pred. No. 0.0016;
Matches 17; Conservative 12; Mismatches 0; Indels 0; Gaps 0;

Qy 1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
:|:||::|:::::|:||||||:||||||:
Db 4 TATGATTCTTTTTGTAAGCCCTAGGGGCT 32

RESULT 3
AAA71131
ID AAA71131 standard; RNA; 42 BP.
XX
AC AAA71131;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #200.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -

XX
PS Example 7; Figure 126; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 9 A; 6 C; 9 G; 18 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
||||||||||||||||||||||||||||||
Db 4 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 32

RESULT 4
AAA70824
ID AAA70824 standard; RNA; 45 BP.
XX
AC AAA70824;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #24.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Homo sapiens.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Claim 220; Page 232; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 45 BP; 11 A; 6 C; 9 G; 19 U; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 UAUGAUUCUUUUUGUAAGCCCUAGGGGC 28
||||||||||||||||||||||||||||
Db 18 UAUGAUUCUUUUUGUAAGCCCUAGGGGC 45

RESULT 5
AAA71087
ID AAA71087 standard; DNA; 46 BP.
XX
AC AAA71087;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #110.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 121; 405pp; English.

XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 11 A; 7 C; 9 G; 19 T; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 60.7%; Pred. No. 0.0045;
Matches 17; Conservative 11; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGC 28
              :|:||::|:::::|:||||||:||||||
Db         19 TATGATTCTTTTTGTAAGCCCTAGGGGC 46
```

RESULT 6
AAA71096
ID AAA71096 standard; DNA; 46 BP.
XX
AC AAA71096;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #119.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 121; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 11 A; 7 C; 9 G; 19 T; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 60.7%; Pred. No. 0.0045;
Matches 17; Conservative 11; Mismatches 0; Indels 0; Gaps 0;

Qy 1 UAUGAUUCUUUUUGUAAGCCCUAGGGGC 28
:|:||::|:::::|:||||||:||||||
Db 19 TATGATTCTTTTTGTAAGCCCTAGGGGC 46

RESULT 7
AAA71099
ID AAA71099 standard; DNA; 46 BP.
XX
AC AAA71099;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #122.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 121; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which

CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 11 A; 7 C; 9 G; 19 T; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 60.7%; Pred. No. 0.0045;
Matches 17; Conservative 11; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGC 28
              :|:||::|:::::|:||||||:||||||
Db         19 TATGATTCTTTTTGTAAGCCCTAGGGGC 46
```

RESULT 8
AAA71100
ID AAA71100 standard; DNA; 46 BP.
XX
AC AAA71100;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #123.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 121; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 11 A; 7 C; 9 G; 19 T; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 60.7%; Pred. No. 0.0045;
Matches 17; Conservative 11; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGC 28
              :|:||::|:::::|:||||||:||||||
Db         19 TATGATTCTTTTTGTAAGCCCTAGGGGC 46
```

RESULT 9
AAA71104
ID AAA71104 standard; RNA; 46 BP.
XX
AC AAA71104;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #180.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of

CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 11 A; 7 C; 9 G; 19 U; 0 other;

```
  Query Match             96.6%;  Score 28;  DB 21;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 0.0045;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGC 28
              ||||||||||||||||||||||||||||
Db         19 UAUGAUUCUUUUUGUAAGCCCUAGGGGC 46
```

RESULT 10
AAA71113
ID AAA71113 standard; RNA; 42 BP.
XX
AC AAA71113;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #189.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 12 A; 7 C; 6 G; 17 U; 0 other;

```
  Query Match             89.0%;  Score 25.8;  DB 21;  Length 42;
  Best Local Similarity   93.1%;  Pred. No. 0.044;
  Matches   27;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
              || |||||||||||||||||||| |||||
Db          4 UAAGAUUCUUUUUGUAAGCCCUACGGGCU 32
```

RESULT 11
AAA71118
ID AAA71118 standard; DNA; 42 BP.
XX
AC AAA71118;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #124.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 125; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual

CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 12 A; 7 C; 6 G; 17 T; 0 other;

Query Match 89.0%; Score 25.8; DB 21; Length 42;
Best Local Similarity 55.2%; Pred. No. 0.044;
Matches 16; Conservative 11; Mismatches 2; Indels 0; Gaps 0;

Qy 1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
:| ||::|:::::|:||||||:| ||||:
Db 4 TAAGATTCTTTTTGTAAGCCCTACGGGCT 32

RESULT 12
AAA71126
ID AAA71126 standard; RNA; 42 BP.
XX
AC AAA71126;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #195.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 126; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 12 A; 7 C; 6 G; 17 U; 0 other;

Query Match 89.0%; Score 25.8; DB 21; Length 42;
Best Local Similarity 93.1%; Pred. No. 0.044;
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
|| |||||||||||||||||||| |||||
Db 4 UAAGAUUCUUUUUGUAAGCCCUACGGGCU 32

RESULT 13
AAA71085
ID AAA71085 standard; DNA; 46 BP.
XX
AC AAA71085;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #108.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 121; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)

CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 12 A; 7 C; 9 G; 18 T; 0 other;

Query Match 85.5%; Score 24.8; DB 21; Length 46;
Best Local Similarity 57.1%; Pred. No. 0.12;
Matches 16; Conservative 10; Mismatches 2; Indels 0; Gaps 0;

Qy 1 UAUGAUUCUUUUUGUAAGCCCUAGGGGC 28
:| ||::|:::::|:||||||:| ||||
Db 19 TAAGATTCTTTTTGTAAGCCCTACGGGC 46

RESULT 14
AAA71103
ID AAA71103 standard; RNA; 46 BP.
XX
AC AAA71103;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #179.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 12 A; 7 C; 9 G; 18 U; 0 other;

Query Match 85.5%; Score 24.8; DB 21; Length 46;
Best Local Similarity 92.9%; Pred. No. 0.12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 UAUGAUUCUUUUUGUAAGCCCUAGGGGC 28
|| |||||||||||||||||||| ||||
Db 19 UAAGAUUCUUUUUGUAAGCCCUACGGGC 46

RESULT 15
AAA71114
ID AAA71114 standard; RNA; 42 BP.
XX
AC AAA71114;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #190.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance

CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 11 A; 8 C; 7 G; 16 U; 0 other;

Query Match 82.1%; Score 23.8; DB 21; Length 42;
Best Local Similarity 92.6%; Pred. No. 0.35;
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGG 27
              || |||||||||||||||||||||| |
Db          4 UAAGAUUCUUUUUGUAAGCCCUAGGCG 30
```

Search completed: January 30, 2004, 08:22:12
Job time : 283.667 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 06:57:47 ; Search time 575.333 Seconds
(without alignments)
2062.073 Million cell updates/sec

Title: US-09-310-844C-24
Perfect score: 29
Sequence: 1 uaugauucuuuuuguaagcccuaggggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 1427288

Minimum DB seq length: 0
Maximum DB seq length: 70

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	15.4	53.1	41	6	AX514720	AX514720 Sequence
c	2	15.4	53.1	41	6	AX520728	AX520728 Sequence
c	3	15.2	52.4	33	6	AR020509	AR020509 Sequence
c	4	15	51.7	40	6	E49126	E49126 Novel G pro
c	5	15	51.7	40	6	E50836	E50836 Novel G pro
c	6	14.8	51.0	35	6	I08597	I08597 Sequence 12
c	7	14.8	51.0	36	6	AR142180	AR142180 Sequence
c	8	14.8	51.0	43	6	AX483394	AX483394 Sequence
	9	14.6	50.3	51	6	AX115517	AX115517 Sequence
	10	14.6	50.3	51	10	AF328713	AF328713 Mus muscu
	11	14.6	50.3	63	9	S63972	S63972 IgH (CDR3 r
c	12	14.6	50.3	65	6	AX484927	AX484927 Sequence
	13	14.4	49.7	25	6	AX042583	AX042583 Sequence
	14	14.4	49.7	25	6	AX043280	AX043280 Sequence
c	15	14.2	49.0	35	6	AX298174	AX298174 Sequence
c	16	14.2	49.0	45	6	I04390	I04390 Sequence 25
c	17	14.2	49.0	56	6	AX247478	AX247478 Sequence
	18	14	48.3	30	6	AR256904	AR256904 Sequence
c	19	14	48.3	30	6	AR256906	AR256906 Sequence
	20	14	48.3	30	6	AX113961	AX113961 Sequence
c	21	14	48.3	30	6	AX113963	AX113963 Sequence
	22	14	48.3	37	6	AR003420	AR003420 Sequence
c	23	14	48.3	37	6	AX555864	AX555864 Sequence
	24	14	48.3	37	6	AX555865	AX555865 Sequence
	25	14	48.3	37	6	I21209	I21209 Sequence 55
	26	14	48.3	37	6	I74476	I74476 Sequence 55
c	27	14	48.3	44	6	AR003419	AR003419 Sequence
c	28	14	48.3	44	6	I21208	I21208 Sequence 54
c	29	14	48.3	44	6	I74475	I74475 Sequence 54
	30	14	48.3	60	6	AX676039	AX676039 Sequence
c	31	14	48.3	62	6	BD034932	BD034932 Sequence
c	32	13.8	47.6	25	6	AX527257	AX527257 Sequence
c	33	13.8	47.6	31	6	AX426018	AX426018 Sequence
c	34	13.8	47.6	36	6	AR142182	AR142182 Sequence
	35	13.8	47.6	42	6	AX017119	AX017119 Sequence
c	36	13.8	47.6	42	6	AX017120	AX017120 Sequence
	37	13.8	47.6	47	6	AR288904	AR288904 Sequence
c	38	13.8	47.6	50	6	AX157596	AX157596 Sequence
c	39	13.8	47.6	50	6	AX164867	AX164867 Sequence
c	40	13.8	47.6	50	6	AX164868	AX164868 Sequence
	41	13.8	47.6	59	10	CRUDELNF	M27128 Chinese Ham
c	42	13.8	47.6	65	6	AX482858	AX482858 Sequence
c	43	13.6	46.9	25	6	AX610026	AX610026 Sequence
c	44	13.6	46.9	25	6	AX610027	AX610027 Sequence
c	45	13.6	46.9	31	6	AX223502	AX223502 Sequence

ALIGNMENTS

RESULT 1
AX514720/c
LOCUS AX514720 41 bp DNA linear PAT 05-OCT-2002
DEFINITION Sequence 918 from Patent WO02052044.
ACCESSION AX514720
VERSION AX514720.1 GI:23561343
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Nakamura,Y., Sekine,A., Iida,A. and Saito,S.
TITLE Detection of genetic polymorphisms
JOURNAL Patent: WO 02052044-A 918 04-JUL-2002;

```
           Riken (JP)
FEATURES             Location/Qualifiers
     source          1. .41
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT       12 a      10 c      10 g       8 t       1 others
ORIGIN

  Query Match             53.1%;  Score 15.4;  DB 6;  Length 41;
  Best Local Similarity   52.0%;  Pred. No. 3.7e+04;
  Matches   13;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCUAGGGGC 28
              ||::|  :::| ||||||:  || |
Db         41 GATTCACTTTGCAAGCCCTCWGGAC 17


RESULT 2
AX520728/c
LOCUS       AX520728                  41 bp    DNA     linear   PAT 05-OCT-2002
DEFINITION  Sequence 6926 from Patent WO02052044.
ACCESSION   AX520728
VERSION     AX520728.1  GI:23571381
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nakamura,Y., Sekine,A., Iida,A. and Saito,S.
  TITLE     Detection of genetic polymorphisms
  JOURNAL   Patent: WO 02052044-A 6926 04-JUL-2002;
            Riken (JP)
FEATURES             Location/Qualifiers
     source          1. .41
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT       12 a      10 c      10 g       8 t       1 others
ORIGIN

  Query Match             53.1%;  Score 15.4;  DB 6;  Length 41;
  Best Local Similarity   52.0%;  Pred. No. 3.7e+04;
  Matches   13;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCUAGGGGC 28
              ||::|  :::| ||||||:  || |
Db         41 GATTCACTTTGCAAGCCCTCWGGAC 17


RESULT 3
AR020509/c
LOCUS       AR020509                  33 bp    DNA     linear   PAT 05-DEC-1998
DEFINITION  Sequence 5 from patent US 5789171.
ACCESSION   AR020509
VERSION     AR020509.1  GI:3975124
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 33)
  AUTHORS   Smeltzer,M.S.
  TITLE     Use of cna, fnba, fnbb, and hlb, gene probes for the
            strain-specific identification of Staphylococcus aureus
  JOURNAL   Patent: US 5789171-A 5 04-AUG-1998;
FEATURES             Location/Qualifiers
     source          1. .33
                     /organism="unknown"
BASE COUNT       12 a       8 c       7 g       6 t
ORIGIN

  Query Match             52.4%;  Score 15.2;  DB 6;  Length 33;
  Best Local Similarity   39.3%;  Pred. No. 4.6e+04;
  Matches   11;  Conservative    9;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          2 AUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
              |:||:: :::: |:||   |   ||||:
Db         32 ATGATTGTTTTAGTAATTTCCCCGGGCT 5


RESULT 4
E49126/c
LOCUS       E49126                    40 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Novel G protein-conjugated receptor protein.
ACCESSION   E49126
VERSION     E49126.1  GI:18629263
KEYWORDS    JP 2001029083-A/4.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   Takasaki,A., Matsumoto,M., Sugimoto,T., Kamahara,M. and Saito,S.
  TITLE     Novel G protein-conjugated receptor protein
  JOURNAL   Patent: JP 2001029083-A 4 06-FEB-2001;
            YAMANOUCHI PHARMACEUT CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 2001029083-A/4
            PD   06-FEB-2001
            PF   23-JUL-1999 JP 1999209918
            PR
            PI   ATSUSHI TAKASAKI,MITSUYUKI MATSUMOTO,TAKASHI SUGIMOTO, PI
            MASAZUMI KAMAHARA,
            PI   SATOSHI SAITO
            PC   C12N15/09,A61K38/00,A61K39/395,A61K39/395,A61K45/00,A61P25/04,
            PC   A61P25/16,
            PC   A61P25/18,C07K14/705,C12N5/10,C12P21/02,C12P21/08,C12Q1/68, PC
              G01N33/15,
            PC   G01N33/50,G01N33/53,G01N33/566,C12N15/00,A61K37/02,C12N5/00 CC

            FH   Key             Location/Qualifiers
            FT   source          1. .40
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT       10 a       8 c       4 g      18 t
ORIGIN

  Query Match             51.7%;  Score 15;  DB 6;  Length 40;
  Best Local Similarity   47.8%;  Pred. No. 5.7e+04;
  Matches   11;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUA 23
              :|:||::|:: : | ||| || |
Db         35 TATGATTCTTATAGAAAGTCCAA 13


RESULT 5
E50836/c
LOCUS       E50836                    40 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Novel G protein-coupled receptor.
ACCESSION   E50836
VERSION     E50836.1  GI:18633541
KEYWORDS    JP 2001054389-A/4.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   Takasaki,A., Matsumoto,M., Sugimoto,T., Kamahara,M. and Saito,S.
  TITLE     Novel G protein-coupled receptor
  JOURNAL   Patent: JP 2001054389-A 4 27-FEB-2001;
```

```
            YAMANOUCHI PHARMACEUT CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 2001054389-A/4
            PD   27-FEB-2001
            PF   17-AUG-1999 JP 1999230777
            PR
            PI   ATSUSHI TAKASAKI,MITSUYUKI MATSUMOTO,TAKASHI SUGIMOTO, PI
            MASAZUMI KAMAHARA,
            PI   SATOSHI SAITO
            PC   C12N15/09,C07K14/705,C07K16/28,C12N1/15,C12N1/19,C12N1/21, PC
             C12N5/10,
            PC   C12P21/02,G01N33/15,G01N33/50//C12P21/08,(C12P21/02,C12R1:91),
            PC   C12N15/00,
            PC   C12N5/00
            CC
            FH   Key             Location/Qualifiers
            FT   source          1. .40
            FT                   /organism='Homo sapiens (human)'.
FEATURES              Location/Qualifiers
     source           1. .40
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
BASE COUNT       10 a       8 c       4 g      18 t
ORIGIN

  Query Match             51.7%;  Score 15;  DB 6;  Length 40;
  Best Local Similarity   47.8%;  Pred. No. 5.7e+04;
  Matches   11;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUA 23
              :|:||::|:: : | ||| || |
Db         35 TATGATTCTTATAGAAAGTCCAA 13


RESULT 6
I08597/c
LOCUS       I08597                   35 bp    DNA     linear   PAT 02-DEC-1994
DEFINITION  Sequence 12 from Patent WO 8707144.
ACCESSION   I08597
VERSION     I08597.1  GI:588701
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 35)
  AUTHORS   Kaufman,R.J., Pittman,D.D. and Toole,J.J.J.
  TITLE     NOVEL PROCOAGULANT PROTEINS
  JOURNAL   Patent: WO 8707144-A 12 03-DEC-1987;
FEATURES              Location/Qualifiers
     source           1. .35
                      /organism="unknown"
BASE COUNT       18 a       8 c       5 g       4 t
ORIGIN

  Query Match             51.0%;  Score 14.8;  DB 6;  Length 35;
  Best Local Similarity   42.3%;  Pred. No. 7.1e+04;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 GAUUCUUUUUGUAAGCCCUAGGGGCU 29
              | ::  ::::| |||| : |||||:
Db         35 GTTTTCTTTTGAAAGCTTTTGGGGCT 10


RESULT 7
AR142180/c
LOCUS       AR142180                 36 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 1 from patent US 6174696.
ACCESSION   AR142180
VERSION     AR142180.1  GI:15102480
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 36)
  AUTHORS   Seman,L.
  TITLE     Method for the determination of homocysteine
  JOURNAL   Patent: US 6174696-A 1 16-JAN-2001;
FEATURES              Location/Qualifiers
     source           1. .36
                      /organism="unknown"
BASE COUNT       12 a       7 c       9 g       8 t
ORIGIN

  Query Match             51.0%;  Score 14.8;  DB 6;  Length 36;
  Best Local Similarity   38.5%;  Pred. No. 7e+04;
  Matches   10;  Conservative    9;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGG 26
              :|: |  |:::::|:  ||| :| ||
Db         33 TATCAAGCTTTTTGTCCGCCATATGG 8


RESULT 8
AX483394/c
LOCUS       AX483394                 43 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 694 from Patent WO02053728.
ACCESSION   AX483394
VERSION     AX483394.1  GI:22317814
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 694 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES              Location/Qualifiers
     source           1. .43
                      /organism="Candida albicans"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:5476"
BASE COUNT       16 a       7 c       4 g      16 t
ORIGIN

  Query Match             51.0%;  Score 14.8;  DB 6;  Length 43;
  Best Local Similarity   38.9%;  Pred. No. 7e+04;
  Matches    7;  Conservative    9;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAG 18
              :|:|| :|:::::|: ||
Db         23 TATGAATCTTTTTGTTAG 6


RESULT 9
AX115517
LOCUS       AX115517                 51 bp    DNA     linear   PAT 11-MAY-2001
DEFINITION  Sequence 640 from Patent WO0129262.
ACCESSION   AX115517
VERSION     AX115517.1  GI:14032459
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Picoult-Newburg,L. and Pohl,M.
  TITLE     Genotyping reagents, kits and methods of use thereof
  JOURNAL   Patent: WO 0129262-A 640 26-APR-2001;
            Orchid BioSciences, Inc. (US)
FEATURES              Location/Qualifiers
     source           1. .51
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:9606"
BASE COUNT       10 a      8 c      8 g     25 t
ORIGIN

  Query Match             50.3%;  Score 14.6;  DB 6;  Length 51;
  Best Local Similarity   31.0%;  Pred. No. 8.5e+04;
  Matches    9;  Conservative   11;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
              :|: |:: ::::: :|   |  || |||:
Db          9 TATTATTTTTTTTATAGCTCACAGTGGCT 37


RESULT 10
AF328713
LOCUS       AF328713                 51 bp    DNA     linear   ROD 29-JUN-2001
DEFINITION  Mus musculus isolate 1.2-11 T-cell receptor beta chain VDJ
            junctional region gene, partial cds.
ACCESSION   AF328713
VERSION     AF328713.1  GI:13898588
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Maryanski,J.L., Attuil,V., Hamrouni,A., Mutin,M., Rossi,M.,
            Aublin,A. and Bucher,P.
  TITLE     Individuality of Ag-selected and preimmune TCR repertoires
  JOURNAL   Immunol. Res. 23 (1), 75-84 (2001)
  MEDLINE   21310440
   PUBMED   11417861
REFERENCE   2  (bases 1 to 51)
  AUTHORS   Maryanski,J.L., Attuil,V., Hamrouni,A., Mutin,M., Rossi,M.,
            Aublin,A. and Bucher,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-2000) INSERM Unit 503, CERVI, 21 Avenue Tony
            Garnier, 69365 Lyon Cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="DBA/2"
                     /isolate="1.2-11"
                     /db_xref="taxon:10090"
                     /rearranged
     mRNA            <1. .>51
                     /product="T-cell receptor beta chain VDJ junctional
                     region"
     CDS             <1. .>51
                     /note="TCRBV10"
                     /codon_start=1
                     /product="T-cell receptor beta chain VDJ junctional
                     region"
                     /protein_id="AAK48813.1"
                     /db_xref="GI:13898589"
                     /translation="SAVYLCASSRGSDYTFG"
BASE COUNT        7 a     16 c     14 g     14 t
ORIGIN

  Query Match             50.3%;  Score 14.6;  DB 10;  Length 51;
  Best Local Similarity   41.4%;  Pred. No. 8.5e+04;
  Matches   12;  Conservative    8;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
              : :|  :|: : :|  |||  :||||||:
Db          6 TGTGTATCTCTGTGCCAGCAGTAGGGGCT 34


RESULT 11
S63972
LOCUS       S63972                   63 bp    DNA     linear   PRI 15-OCT-1993
DEFINITION  IgH {CDR3 region, VHDJH rearrangement, clone DV3-R2} [human,
            leukemia patient D, bone marrow B-cells, Genomic, 63 nt].
ACCESSION   S63972
VERSION     S63972.1  GI:408181
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 63)
  AUTHORS   Steenbergen,E.J., Verhagen,O.J., van Leeuwen,E.F., von dem
            Borne,A.E. and van der Schoot,C.E.
  TITLE     Distinct ongoing Ig heavy chain rearrangement processes in
            childhood B-precursor acute lymphoblastic leukemia
  JOURNAL   Blood 82 (2), 581-589 (1993)
  MEDLINE   93320440
   PUBMED   8329713
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 135423] from the original journal article.
            This sequence comes from Fig. 3D.
FEATURES             Location/Qualifiers
     source          1. .63
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     gene            1. .63
                     /partial
                     /gene="IgH"
BASE COUNT       16 a     11 c     19 g     17 t
ORIGIN

  Query Match             50.3%;  Score 14.6;  DB 9;  Length 63;
  Best Local Similarity   41.4%;  Pred. No. 8.4e+04;
  Matches   12;  Conservative    8;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
              :|:||:  :  : |: |   |:||||||:
Db         17 TATGATAGTAGTGGTTATTACTAGGGGCT 45


RESULT 12
AX484927/c
LOCUS       AX484927                 65 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 2227 from Patent WO02053728.
ACCESSION   AX484927
VERSION     AX484927.1  GI:22319211
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 2227 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .65
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5476"
BASE COUNT       22 a      9 c      7 g     27 t
ORIGIN

  Query Match             50.3%;  Score 14.6;  DB 6;  Length 65;
  Best Local Similarity   27.6%;  Pred. No. 8.4e+04;
  Matches    8;  Conservative   12;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 UAUGAUUCUUUUUGUAAGCCCUAGGGGCU 29
              :|:  :: ::::::|:||   |:||   |:
Db         57 TATTTTTTTTTTTGTAAAGACTAGAACCT 29


RESULT 13
```

```
AX042583
LOCUS       AX042583                  25 bp    DNA     linear   PAT 23-NOV-2000
DEFINITION  Sequence 149 from Patent WO0065088.
ACCESSION   AX042583
VERSION     AX042583.1  GI:11341191
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Ulfendahl,P.J. and Wong,K.C.
  TITLE     Primers for identifying typing or classifying nucleic acids
  JOURNAL   Patent: WO 0065088-A 149 02-NOV-2000;
            Amersham Pharmacia Biotech AB (SE)
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="DQA1 Homozygote primer sequence"
BASE COUNT        3 a      3 c      5 g     14 t
ORIGIN

  Query Match             49.7%;  Score 14.4;  DB 6;  Length 25;
  Best Local Similarity   33.3%;  Pred. No. 1.1e+05;
  Matches    8;  Conservative   10;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          6 UUCUUUUUGUAAGCCCUAGGGGCU 29
              :: ::::: : ||| |: ||| |:
Db          2 TTTTTTTTTTTAGCACTGGGGACT 25


RESULT 14
AX043280
LOCUS       AX043280                  25 bp    DNA     linear   PAT 23-NOV-2000
DEFINITION  Sequence 846 from Patent WO0065088.
ACCESSION   AX043280
VERSION     AX043280.1  GI:11341888
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Ulfendahl,P.J. and Wong,K.C.
  TITLE     Primers for identifying typing or classifying nucleic acids
  JOURNAL   Patent: WO 0065088-A 846 02-NOV-2000;
            Amersham Pharmacia Biotech AB (SE)
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="DQA1 Heterozygote Primer Sequence"
BASE COUNT        3 a      3 c      5 g     14 t
ORIGIN

  Query Match             49.7%;  Score 14.4;  DB 6;  Length 25;
  Best Local Similarity   33.3%;  Pred. No. 1.1e+05;
  Matches    8;  Conservative   10;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          6 UUCUUUUUGUAAGCCCUAGGGGCU 29
              :: ::::: : ||| |: ||| |:
Db          2 TTTTTTTTTTTAGCACTGGGGACT 25


RESULT 15
AX298174/c
LOCUS       AX298174                  35 bp    DNA     linear   PAT 26-NOV-2001
DEFINITION  Sequence 20 from Patent WO0183788.
ACCESSION   AX298174
VERSION     AX298174.1  GI:17128241
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Heitz,T., Dhondt,S., Geoffroy,P., Legrand,M. and Gouzerh,G.
  TITLE     Plant pla2 polypeptides involved in plant defence reaction,
            polyncleotides encoding said polypeptides and transformed plants
            containing them
  JOURNAL   Patent: WO 0183788-A 20 08-NOV-2001;
            Rhobio (FR)
FEATURES             Location/Qualifiers
     source          1. .35
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Oligonucleotide amorce NtPat3pGEX sens"
BASE COUNT       14 a      5 c     10 g      6 t
ORIGIN

  Query Match             49.0%;  Score 14.2;  DB 6;  Length 35;
  Best Local Similarity   40.7%;  Pred. No. 1.3e+05;
  Matches   11;  Conservative    8;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          2 AUGAUUCUUUUUGUAAGCCCUAGGGGC 28
              |:  ::| ::: |:|| | |:|| | |
Db         27 ATCTTTCCTTTGGTAACCTCTAGAGTC 1


Search completed: January 30, 2004, 08:51:23
Job time : 579.333 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 07:55:03 ; Search time 1615 Seconds
(without alignments)
436.427 Million cell updates/sec

Title: US-09-310-844C-23
Perfect score: 29
Sequence: 1 nnngauncuuunnguaagcccnangngnn 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 243536

Minimum DB seq length: 0
Maximum DB seq length: 70

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	13	44.8	46	28	AZ833686	AZ833686 2M0115L20
	2	12.8	44.1	70	28	BH759592	BH759592 KG05236-3
c	3	12.4	42.8	52	9	AA700959	AA700959 zf87d10.s
c	4	12.4	42.8	70	9	AA468615	AA468615 ne08b04.s
	5	12.2	42.1	48	28	AZ503560	AZ503560 1M0343E21
	6	12.2	42.1	56	29	BZ770420	BZ770420 SALK_1433
c	7	11.8	40.7	40	9	AA975071	AA975071 on03d07.s
c	8	11.8	40.7	49	10	BE970036	BE970036 601680150
c	9	11.8	40.7	65	9	AA733449	AA733449 vt73h08.r
	10	11.8	40.7	70	9	AI767928	AI767928 wi99c01.x
c	11	11.8	40.7	70	28	BH216023	BH216023 1006039G0
	12	11.6	40.0	49	14	U44334	U44334 ENU44334 As
c	13	11.6	40.0	58	9	AI584456	AI584456 fb93h12.x
	14	11.6	40.0	70	9	AI814489	AI814489 wj73g11.x
c	15	11.4	39.3	32	29	HSMC42B09	X88068 H.sapiens D
	16	11.4	39.3	34	29	BX001854	BX001854 Arabidops
c	17	11.4	39.3	40	29	TA253H01Q	AL483109 T. brucei
	18	11.4	39.3	46	9	AI887082	AI887082 wl96e09.x
c	19	11.4	39.3	66	9	AA247859	AA247859 j3371.seq
	20	11.4	39.3	67	29	TA113E04Q	AL460350 T. brucei
	21	11.4	39.3	68	9	AU254479	AU254479 AU254479
	22	11.2	38.6	22	14	D18745	D18745 MUSGS01807
	23	11.2	38.6	30	9	AU259912	AU259912 AU259912
c	24	11.2	38.6	37	28	AZ950243	AZ950243 2M0214C15
	25	11.2	38.6	44	29	AL771575	AL771575 Arabidops
c	26	11.2	38.6	46	9	AA522160	AA522160 vf97c10.r
	27	11.2	38.6	49	9	AV841468	AV841468 AV841468
	28	11.2	38.6	54	28	BH224605	BH224605 1006120A0
	29	11.2	38.6	60	14	H19786	H19786 yn60a10.s1
	30	11.2	38.6	64	9	AI321110	AI321110 d4c09nm.r
c	31	11.2	38.6	65	29	AL755798	AL755798 Arabidops
	32	11.2	38.6	66	9	AI571487	AI571487 tr56g10.x
c	33	11.2	38.6	67	9	AA936041	AA936041 nz53f10.s
	34	11.2	38.6	70	9	AU258214	AU258214 AU258214
c	35	11.2	38.6	70	14	CB916098	CB916098 VVD106F01
c	36	11	37.9	34	29	TA98E04P	AL461291 T. brucei
	37	11	37.9	40	9	AI016485	AI016485 ot86g10.s
	38	11	37.9	43	28	AZ639598	AZ639598 1M0501L07
c	39	11	37.9	56	12	BM445434	BM445434 1IL11C9.a
c	40	10.8	37.2	31	9	AA206191	AA206191 zq54a06.r
c	41	10.8	37.2	33	9	AU256066	AU256066 AU256066
c	42	10.8	37.2	44	28	BH811570	BH811570 SALK_0591
	43	10.8	37.2	45	29	BX289611	BX289611 Arabidops
	44	10.8	37.2	49	9	AI273069	AI273069 qv62g02.x
c	45	10.8	37.2	50	28	AZ817325	AZ817325 2M0086C23

ALIGNMENTS

RESULT 1
AZ833686
LOCUS AZ833686 46 bp DNA linear GSS 20-FEB-2001
DEFINITION 2M0115L20R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0115L20 R, genomic survey sequence.
ACCESSION AZ833686
VERSION AZ833686.1 GI:13003594
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 46)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0115 row: L column: 20
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop:·46.

FEATURES Location/Qualifiers
source 1. .46
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0115L20"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 14 a 6 c 13 g 13 t
ORIGIN

Query Match 44.8%; Score 13; DB 28; Length 46;
Best Local Similarity 50.0%; Pred. No. 1.5e+04;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 4 GAUNCUUUNNGUAAGC 19
||: |::: |:||||
Db 8 GATACTTTAAGTAAGC 23

RESULT 2
BH759592
LOCUS BH759592 70 bp DNA linear GSS 12-MAR-2002
DEFINITION KG05236-3prime Drosophila melanogaster P{SUPor-P} P element insertion lines Drosophila melanogaster genomic Sequence recovered from 3' end of P element, genomic survey sequence.
ACCESSION BH759592
VERSION BH759592.1 GI:19352831
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 70)
AUTHORS Levis,R., Hoskins,R., Liao,G., Mozden,N., Tsang,G., He,Y., Karpen,G., Bellen,H., Rubin,G. and Spradling,A.
TITLE The Berkeley Drosophila Genome Project Gene Disruption Project
JOURNAL Unpublished
COMMENT Contact: Gerald Rubin
Berkeley Drosophila Genome Project
University of California, Berkeley
LSA Building, Berkeley, CA 94720-3200, USA
Fax: 5106439947
Email: gerry@fruitfly.berkeley.edu
Sequence recovery method was inverse PCR.
Sequence orientation is forward strand relative to 5' end of P element
The P element insertion position is base 1 in the 70 bases. This insertion position refers to the first base of the 8 base target recognition sequence.
Class: transposon-tagged.

FEATURES Location/Qualifiers
source 1. .70
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_lib="Drosophila melanogaster P{SUPor-P} P element insertion lines"
/note="Inverse PCR was performed on Drosophila melanogaster strains each of which contains one or more P{SUPor-P} P-element transposon insertion. The resultant fragment for each strain was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://www.fruitfly.org/about/methods/inverse.pcr.html."

BASE COUNT 20 a 11 c 9 g 30 t
ORIGIN

Query Match 44.1%; Score 12.8; DB 28; Length 70;
Best Local Similarity 42.9%; Pred. No. 2.1e+04;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 5 AUNCUUUNNGUAAGCCCNANG 25
|: |::: :|| ||| | |
Db 15 ATACTTTATTTAATCCCAAAG 35

RESULT 3
AA700959/c
LOCUS AA700959 52 bp mRNA linear EST 19-DEC-1997
DEFINITION zf87d10.s1 Soares_pineal_gland_N3HPG Homo sapiens cDNA clone IMAGE:383923 3' similar to TR:P79324 P79324 RIBOSOMAL PROTEIN L15 ;, mRNA sequence.
ACCESSION AA700959
VERSION AA700959.1 GI:2704124
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 52)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.

FEATURES Location/Qualifiers
source 1. .52
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:1292180"
/db_xref="taxon:9606"
/clone="IMAGE:383923"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares_pineal_gland_N3HPG"

/note="Organ: pineal gland; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCGTTTTTTTTTTTTTTTTT 3'] , double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo. "

BASE COUNT 17 a 12 c 12 g 11 t
ORIGIN

Query Match 42.8%; Score 12.4; DB 9; Length 52;
Best Local Similarity 52.9%; Pred. No. 3.2e+04;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```
Qy          4 GAUNCUUUNNGUAAGCC 20
              ||: | ::  |:|||||
Db         31 GATCCATTCCGTAAGCC 15
```

RESULT 4
AA468615/c
LOCUS AA468615 70 bp mRNA linear EST 15-AUG-1997
DEFINITION ne08b04.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:880591 3' similar to TR:G178246 G178246 MATURE ALPHA-GALACTOSIDASE A, COMPLETE CDS PRECURSOR ;, mRNA sequence.
ACCESSION AA468615
VERSION AA468615.1 GI:2195149
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 70)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D. , Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1285 Std Error: 0.00
Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1. .70
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:880591"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Co3"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. "
BASE COUNT 17 a 22 c 14 g 17 t
ORIGIN

Query Match 42.8%; Score 12.4; DB 9; Length 70;
Best Local Similarity 52.9%; Pred. No. 3.5e+04;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```
Qy          4 GAUNCUUUNNGUAAGCC 20
              ||: |::  | |||||
Db         64 GATCCTTTGGGCAAGCC 48
```

RESULT 5
AZ503560
LOCUS AZ503560 48 bp DNA linear GSS 05-OCT-2000
DEFINITION 1M0343E21F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0343E21 F, genomic survey sequence.
ACCESSION AZ503560
VERSION AZ503560.1 GI:10684876
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 48)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0343 row: E column: 21
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 48.
FEATURES Location/Qualifiers
source 1. .48
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0343E21"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 14 a 7 c 17 g 10 t
ORIGIN

```
  Query Match           42.1%;  Score 12.2;  DB 28;  Length 48;
  Best Local Similarity  43.5%;  Pred. No. 4e+04;
  Matches   10;  Conservative    4;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          5 AUNCUUUNNGUAAGCCCNANGNG 27
              |: |: :  |:||| |  | | |
Db         13 ATGCTCTGGGTAAGGCACAAGAG 35


RESULT 6
BZ770420
LOCUS       BZ770420                 56 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  SALK_143355.56.00.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_143355.56.00.x, genomic
            survey sequence.
ACCESSION   BZ770420
VERSION     BZ770420.1  GI:28944104
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 56)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA. This sequence lies within an annotated exon of At5g40030.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .56
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_143355.56.00.x"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
BASE COUNT       14 a     12 c     10 g     20 t
ORIGIN

  Query Match           42.1%;  Score 12.2;  DB 29;  Length 56;
  Best Local Similarity  45.5%;  Pred. No. 4.2e+04;
  Matches   10;  Conservative    4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANG 25
              ||: |: :   :|||||  | |
Db          3 GATACTATTGTTAAGCCTAACG 24


RESULT 7
AA975071/c
LOCUS       AA975071                 40 bp    mRNA    linear   EST 26-AUG-1998
DEFINITION  on03d07.s1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1555597 3'
            similar to TR:P70566 P70566 N-TROPOMODULIN. ;, mRNA sequence.
ACCESSION   AA975071
VERSION     AA975071.1  GI:3150863
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 2096   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1555597"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Kid3"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer,
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia),  digested with Not I and cloned into the Not
                     I and Eco RI sites  of the modified pT7T3 vector.  mRNA
                     source: 2 pooled kidneys. Library went through one round
                     of normalization. Library constructed by Bento Soares and
                     M. Fatima Bonaldo.  "
BASE COUNT       10 a      7 c     11 g     12 t
ORIGIN

  Query Match           40.7%;  Score 11.8;  DB 9;  Length 40;
  Best Local Similarity  40.0%;  Pred. No. 6.2e+04;
  Matches    8;  Conservative    5;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          6 UNCUUUNNGUAAGCCCNANG 25
              : |:::  |:||| ||   |
Db         35 TCCTTTCCGTAAGACCTTGG 16


RESULT 8
BE970036/c
LOCUS       BE970036                 49 bp    mRNA    linear   EST 04-OCT-2000
DEFINITION  601680150F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3950172 5',
            mRNA sequence.
ACCESSION   BE970036
VERSION     BE970036.1  GI:10582969
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM816 row: d column: 13
High quality sequence stop: 49.

FEATURES Location/Qualifiers
source 1. .49
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3950172"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_78"
/note="Organ: pancreas; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.2 kb (range 0.5-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 20 a 7 c 12 g 10 t
ORIGIN

Query Match 40.7%; Score 11.8; DB 10; Length 49;
Best Local Similarity 45.0%; Pred. No. 6.6e+04;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

```
Qy          6 UNCUUUNNGUAAGCCCNANG 25
              :  :::  | ||||| | |
Db         48 TTTTTTATGCAAGCCCCAGG 29
```

RESULT 9
AA733449/c
LOCUS AA733449 65 bp mRNA linear EST 07-JAN-1998
DEFINITION vt73h08.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA clone IMAGE:1176831 5' similar to gb:X06617 40S RIBOSOMAL PROTEIN S11 (HUMAN);, mRNA sequence.
ACCESSION AA733449
VERSION AA733449.1 GI:2755116
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 65)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:634679
Trace considered overall poor quality
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 1.

FEATURES Location/Qualifiers
source 1. .65
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:1176831"
/dev_stage="8 weeks"
/lab_host="DH10B"
/clone_lib="Barstead mouse irradiated colon MPLRB7"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained from 8 week old mouse. Colon was harvested 72 hours after irradiation with 1400 Gys. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATTCGGATCCTTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead. "

BASE COUNT 23 a 15 c 16 g 11 t
ORIGIN

Query Match 40.7%; Score 11.8; DB 9; Length 65;
Best Local Similarity 44.4%; Pred. No. 7.2e+04;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

```
Qy          4 GAUNCUUUNNGUAAGCCC 21
              | : |:::  |:|||| |
Db         44 GCTGCTTTTGGTAAGCAC 27
```

RESULT 10
AI767928
LOCUS AI767928 70 bp mRNA linear EST 21-DEC-1999
DEFINITION wi99c01.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2401440 3' similar to SW:RT14_HUMAN P78537 RT14 PROTEIN ;, mRNA sequence.
ACCESSION AI767928
VERSION AI767928.1 GI:5234426
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 70)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert Length: 574 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.

FEATURES Location/Qualifiers
source 1. .70
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2401440"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Kid12"

/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_Kid5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1323912-1325831, 1471368-1472903 and 1492104-1493255). Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 20 a 24 c 6 g 20 t
ORIGIN

Query Match 40.7%; Score 11.8; DB 9; Length 70;
Best Local Similarity 47.4%; Pred. No. 7.3e+04;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 5 AUNCUUUNNGUAAGCCCNA 23
Db 46 ATTCTTTAAGCAAGCCAGA 64

RESULT 11
BH216023/c
LOCUS BH216023 70 bp DNA linear GSS 08-NOV-2001
DEFINITION 1006039G04.2EL_y1 1006 - RescueMu Grid G Zea mays genomic, genomic survey sequence.
ACCESSION BH216023
VERSION BH216023.1 GI:16806681
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 70)
AUTHORS Walbot,V.
TITLE Maize genomic sequences found using engineered RescueMu transposon
JOURNAL Unpublished
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1006039 row: 43
Class: transposon-tagged.
FEATURES Location/Qualifiers
source 1. .70
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="1006 - RescueMu Grid G"
/note="Organ: leaf; Vector: RescueMu (engineered from pBlueScript backbone); Site_1: BamHI; Site_2: BglII; RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'RescueMu.' Grid G was grown at Stanford in 2000. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."
BASE COUNT 20 a 16 c 18 g 16 t
ORIGIN

Query Match 40.7%; Score 11.8; DB 28; Length 70;
Best Local Similarity 50.0%; Pred. No. 7.3e+04;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 4 GAUNCUUUNNGUAAGCCC 21
Db 70 GATCCTTTTAGGAGGCCC 53

RESULT 12
U44334
LOCUS U44334 49 bp mRNA linear EST 03-APR-1996
DEFINITION ENU44334 Aspergillus nidulans cleistothecium Emericella nidulans cDNA clone SE0762, mRNA sequence.
ACCESSION U44334
VERSION U44334.1 GI:1244997
KEYWORDS EST.
SOURCE Emericella nidulans (anamorph: Aspergillus nidulans)
ORGANISM Emericella nidulans
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; Emericella.
REFERENCE 1 (bases 1 to 49)
AUTHORS Lee,D., Lee,S., Hwang,H., Kim,J. and Chae,K.
TITLE Quantitative analysis of gene expression in sexual structures of Aspergillus nidulans by sequencing of 3'-directed cDNA clones
JOURNAL FEMS Microbiol. Lett. 138 (1), 71-76 (1996)
MEDLINE 96236220
PUBMED 8674973
COMMENT Contact: Keon-Sang Chae
Chonbuk National University
Chonju, 561-756, S. Korea
Tel: +82-652-70-3340
Fax: +82-652-70-3345
Email: chaeks@chonbuknms.chonbuk.ac.kr.
FEATURES Location/Qualifiers
source 1. .49
/organism="Emericella nidulans"
/mol_type="mRNA"
/strain="FGSC4"
/db_xref="taxon:162425"
/clone="SE0762"
/tissue_type="cleistothecium"
/cell_type="Hull cell"
/dev_stage="sexual"
/clone_lib="Aspergillus nidulans cleistothecium"
/note="3'-directed cDNA clones; single-pass sequencing"
BASE COUNT 11 a 13 c 10 g 15 t
ORIGIN

Query Match 40.0%; Score 11.6; DB 14; Length 49;
Best Local Similarity 37.5%; Pred. No. 8.4e+04;
Matches 9; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Qy 4 GAUNCUUUNNGUAAGCCCNANGNG 27
Db 21 GATTCTTTCATTAACTCCCAACGG 44

RESULT 13
AI584456/c
LOCUS AI584456 58 bp mRNA linear EST 07-JUN-2001
DEFINITION fb93h12.x1 Zebrafish WashU MPIMG EST Danio rerio cDNA clone IMAGE:3719495 3' similar to SW:TRF1_SALSA P80426 SEROTRANSFERRIN I PRECURSOR ;, mRNA sequence.
ACCESSION AI584456
VERSION AI584456.1 GI:4570353
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes

```
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 58)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Seq primer: T7 ET from Amersham
            High quality sequence stop: 1
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .58
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="IMAGE:3719495"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT3']
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries.  Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT       19 a      17 c       9 g      13 t
ORIGIN

  Query Match             40.0%;  Score 11.6;  DB 9;  Length 58;
  Best Local Similarity   37.5%;  Pred. No. 8.9e+04;
  Matches    9;  Conservative    5;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              | : |:::  :|||| |   | |
Db         25 GTTGCTTTTATTAAGCACACTGTG 2


RESULT 14
AI814489
LOCUS       AI814489                 70 bp    mRNA    linear   EST 24-AUG-1999
DEFINITION  wj73g11.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2408516 3'
            similar to gb:X59268 TRANSCRIPTION INITIATION FACTOR IIB (HUMAN);,
            mRNA sequence.
ACCESSION   AI814489
VERSION     AI814489.1  GI:5425704
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 70)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .70
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2408516"
                     /tissue_type="squamous cell carcinoma, poorly
                     differentiated (4 pooled tumors, including primary and
                     metastatic)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NCI_CGAP_Lu19"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     pooled lung tumor tissue, and was then primed with a Not I
                     - oligo(dT) primer. Double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I  and Eco RI sites of the modified
                     pT7T3 vector.  Library went througb one round of
                     normalization. Library constructed by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT       17 a      5 c      32 g      16 t
ORIGIN

  Query Match             40.0%;  Score 11.6;  DB 9;  Length 70;
  Best Local Similarity   45.8%;  Pred. No. 9.4e+04;
  Matches   11;  Conservative    3;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              |:  |: :  |:|||  | | | |
Db         24 GAGTCTCTTGGTAAGGGCGATGAG 47


RESULT 15
HSMC42B09/c
LOCUS       HSMC42B09                32 bp    DNA     linear   GSS 29-MAY-1997
DEFINITION  H.sapiens DNA for trapped exon (ID HMC42B09), genomic survey
            sequence.
ACCESSION   X88068
VERSION     X88068.1  GI:1437990
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 32)
  AUTHORS   Chen,H., Chrast,R., Rossier,C., Morris,M.A., Lalioti,M.D. and
            Antonarakis,S.E.
  TITLE     Cloning of 559 potential exons of genes of human chromosome 21 by
            exon trapping
  JOURNAL   Genome Res. 6 (8), 747-760 (1996)
  MEDLINE   97011340
   PUBMED   8858350
REFERENCE   2  (bases 1 to 32)
  AUTHORS   Chen,H.M., Rossier,C., Chrast,R. and Antonarakis,S.E.
  TITLE     Cloning of trapped exons from human chromosome 21
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 32)
  AUTHORS   Antonarakis,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1995) Stylianos E. Antonarakis, Division of
            Medical Genetics, University and Cantonal Hospital of Geneva, CMU,
            1 rue Michel-Servet, 1211 Geneva, SWITZERLAND
FEATURES             Location/Qualifiers
     source          1. .32
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
     exon            1. .32
                     /note="trapped exon"
BASE COUNT        6 a     11 c      5 g      9 t      1 others
ORIGIN

  Query Match             39.3%;  Score 11.4;  DB 29;  Length 32;
  Best Local Similarity   47.1%;  Pred. No. 9.6e+04;
  Matches    8;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCC 20
              ||: |:::    |||||
Db         31 GATACTTTCANCAAGCC 15


Search completed: January 30, 2004, 10:12:29
Job time : 1627 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 08:51:33 ; Search time 176 Seconds
(without alignments)
600.524 Million cell updates/sec

Title: US-09-310-844C-23
Perfect score: 29
Sequence: 1 nnngauncuuunnguaagcccnangngnn 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 1462844

Minimum DB seq length: 0
Maximum DB seq length: 70

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	12.8	44.1	68	8	US-08-781-986A-2762	Sequence 2762, Ap
c	2	12.4	42.8	60	13	US-09-908-975-18725	Sequence 18725, A
c	3	12.4	42.8	65	13	US-09-908-975-2848	Sequence 2848, Ap
	4	12.2	42.1	50	12	US-10-131-827-464	Sequence 464, App
	5	12.2	42.1	60	13	US-09-908-975-9828	Sequence 9828, Ap
c	6	12.2	42.1	60	13	US-09-908-975-12187	Sequence 12187, A
c	7	12.2	42.1	60	13	US-09-908-975-15109	Sequence 15109, A
	8	12.2	42.1	60	13	US-09-908-975-18934	Sequence 18934, A
	9	11.8	40.7	25	15	US-10-098-263B-76444	Sequence 76444, A
	10	11.8	40.7	47	10	US-09-230-926A-35	Sequence 35, Appl
c	11	11.8	40.7	60	13	US-09-908-975-15914	Sequence 15914, A
	12	11.8	40.7	60	13	US-09-908-975-17626	Sequence 17626, A
	13	11.8	40.7	65	13	US-09-908-975-1254	Sequence 1254, Ap
	14	11.8	40.7	65	13	US-09-908-975-30297	Sequence 30297, A
	15	11.8	40.7	65	13	US-10-032-585-316	Sequence 316, App
	16	11.6	40.0	24	11	US-09-964-895-27	Sequence 27, Appl
	17	11.6	40.0	24	13	US-10-059-152-26	Sequence 26, Appl
c	18	11.6	40.0	31	11	US-09-848-754A-6937	Sequence 6937, Ap
c	19	11.6	40.0	31	11	US-09-848-754A-7188	Sequence 7188, Ap
c	20	11.6	40.0	31	11	US-09-848-754A-7495	Sequence 7495, Ap
c	21	11.6	40.0	31	11	US-09-740-332-6639	Sequence 6639, Ap
c	22	11.6	40.0	31	11	US-09-740-332-9154	Sequence 9154, Ap
c	23	11.6	40.0	31	13	US-09-817-879-6639	Sequence 6639, Ap
c	24	11.6	40.0	31	13	US-09-817-879-9154	Sequence 9154, Ap
c	25	11.6	40.0	31	15	US-10-163-552-1019	Sequence 1019, Ap
c	26	11.6	40.0	31	15	US-10-156-306-3281	Sequence 3281, Ap
c	27	11.6	40.0	36	9	US-09-904-599A-7	Sequence 7, Appli
	28	11.6	40.0	38	12	US-10-388-329-15	Sequence 15, Appl
c	29	11.6	40.0	56	13	US-09-800-130A-8	Sequence 8, Appli
c	30	11.6	40.0	56	13	US-10-413-909-8	Sequence 8, Appli
	31	11.6	40.0	60	13	US-09-908-975-5781	Sequence 5781, Ap
c	32	11.6	40.0	60	13	US-09-908-975-12753	Sequence 12753, A
c	33	11.6	40.0	60	13	US-09-908-975-14781	Sequence 14781, A
	34	11.6	40.0	65	13	US-09-908-975-24835	Sequence 24835, A
	35	11.6	40.0	65	13	US-09-908-975-29918	Sequence 29918, A
c	36	11.4	39.3	25	15	US-10-098-263B-5191	Sequence 5191, Ap
c	37	11.4	39.3	25	15	US-10-098-263B-5192	Sequence 5192, Ap
	38	11.4	39.3	25	15	US-10-098-263B-94421	Sequence 94421, A
	39	11.4	39.3	44	13	US-10-053-530-28	Sequence 28, Appl
	40	11.4	39.3	44	15	US-10-207-655-28	Sequence 28, Appl
	41	11.4	39.3	60	13	US-09-908-975-6202	Sequence 6202, Ap
c	42	11.4	39.3	60	13	US-09-908-975-7920	Sequence 7920, Ap
c	43	11.4	39.3	60	13	US-09-908-975-20533	Sequence 20533, A
	44	11.4	39.3	65	13	US-09-908-975-4740	Sequence 4740, Ap
c	45	11.2	38.6	25	12	US-10-346-880-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-08-781-986A-2762
; Sequence 2762, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 2762:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 68 base pairs
; TYPE: nucleic acid

```
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-781-986A-2762

  Query Match             44.1%;  Score 12.8;  DB 8;  Length 68;
  Best Local Similarity   47.6%;  Pred. No. 2.1e+03;
  Matches   10;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          5 AUNCUUUNNGUAAGCCCNANG 25
              |: |: :   :||||| | |
Db          3 ATCCTGTTCCTAAGCCCGACG 23


RESULT 2
US-09-908-975-18725/c
; Sequence 18725, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18725
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-908-975-18725

  Query Match             42.8%;  Score 12.4;  DB 13;  Length 60;
  Best Local Similarity   55.6%;  Pred. No. 3.5e+03;
  Matches   10;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          8 CUUUNNGUAAGCCCNANG 25
              |:::  | ||||| | |
Db         25 CTTTCTGAAAGCCCCATG 8


RESULT 3
US-09-908-975-2848/c
; Sequence 2848, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2848
;   LENGTH: 65
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-09-908-975-2848

  Query Match             42.8%;  Score 12.4;  DB 13;  Length 65;
  Best Local Similarity   52.9%;  Pred. No. 3.6e+03;
  Matches    9;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCC 20
              ||: |::   |:|||||
Db         53 GATACTTGCAGTAAGCC 37


RESULT 4
US-10-131-827-464
; Sequence 464, Application US/10131827
; Publication No. US20040009479A1
; GENERAL INFORMATION:
;  APPLICANT: Wohlgemuth, Jay
;  APPLICANT:  Fry, Kirk
;  APPLICANT:  Woodward, Robert
;  APPLICANT:  Ly, Ngoc
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND MONITORING AUTOIMMUNE
;  TITLE OF INVENTION:  CHRONIC INFLAMMATORY DISEASES
;  FILE REFERENCE: 506612000120
;  CURRENT APPLICATION NUMBER: US/10/131,827
;  CURRENT FILING DATE:  2002-09-06
;  PRIOR APPLICATION NUMBER: US 10/006,290
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: US 60/296,764
;  PRIOR FILING DATE: 2001-06-08
;  NUMBER OF SEQ ID NOS: 9090
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 464
;   LENGTH: 50
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-131-827-464

  Query Match             42.1%;  Score 12.2;  DB 12;  Length 50;
  Best Local Similarity   45.5%;  Pred. No. 4.4e+03;
  Matches   10;  Conservative    4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANG 25
              ||  |:::   :||| || | |
Db         22 GAGGCTTTTCCTAAGGCCAAGG 43


RESULT 5
US-09-908-975-9828
; Sequence 9828, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE \
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9828
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-908-975-9828
```

```
  Query Match             42.1%;  Score 12.2;  DB 13;  Length 60;
  Best Local Similarity   52.2%;  Pred. No. 4.6e+03;
  Matches   12;  Conservative    2;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          5 AUNCUUUNNGUAAGCCCNANGNG 27
              |  |:    |:||||| | | |
Db         34 AACCTCATGGTAAGCCCAACGTG 56


RESULT 6
US-09-908-975-12187/c
; Sequence 12187, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12187
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-908-975-12187

  Query Match             42.1%;  Score 12.2;  DB 13;  Length 60;
  Best Local Similarity   40.9%;  Pred. No. 4.6e+03;
  Matches    9;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANG 25
              ||: |:::  |:|  | | | |
Db         54 GATTCTTTCTGTAGCCGCTAAG 33


RESULT 7
US-09-908-975-15109/c
; Sequence 15109, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15109
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-908-975-15109

  Query Match             42.1%;  Score 12.2;  DB 13;  Length 60;
  Best Local Similarity   45.5%;  Pred. No. 4.6e+03;
  Matches   10;  Conservative    4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANG 25
              ||: |:::  | | | || | |
Db         37 GATGCTTTCTGCATGGCCAAGG 16


RESULT 8
US-09-908-975-18934
; Sequence 18934, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE \
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18934
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-908-975-18934

  Query Match             42.1%;  Score 12.2;  DB 13;  Length 60;
  Best Local Similarity   43.5%;  Pred. No. 4.6e+03;
  Matches   10;  Conservative    4;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          5 AUNCUUUNNGUAAGCCCNANGNG 27
              |: | ::  |:|||  |   | |
Db         27 ATCCGTTGCGTAAGCACAGAGAG 49


RESULT 9
US-10-098-263B-76444
; Sequence 76444, Application US/10098263B
; Publication No. US20030104410A1
; GENERAL INFORMATION:
;  APPLICANT: Mittman, Michael
;  TITLE OF INVENTION: Human Microarray
;  FILE REFERENCE: 3118.1
;  CURRENT APPLICATION NUMBER: US/10/098,263B
;  CURRENT FILING DATE:  2003-01-08
;  PRIOR APPLICATION NUMBER: 60/276,759
;  PRIOR FILING DATE: 2001-03-16
;  NUMBER OF SEQ ID NOS: 131066
;  SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 76444
;   LENGTH: 25
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-098-263B-76444

  Query Match             40.7%;  Score 11.8;  DB 15;  Length 25;
  Best Local Similarity   44.4%;  Pred. No. 6.7e+03;
  Matches    8;  Conservative    5;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCC 21
              ||: |:::   :||| ||
Db          1 GATACTTTTTTTAAGTCC 18
```

```
RESULT 10
US-09-230-926A-35
; Sequence 35, Application US/09230926A
; Patent No. US20020168633A1
; GENERAL INFORMATION:
;  APPLICANT: MABILAT, Claude
;  APPLICANT:  SCHLEIFER, Karl-Heinz
;  APPLICANT:  LUDWIG, Wolfgang
;  TITLE OF INVENTION: NUCLEOTIDE FRAGMENT OF THE 23S RNA OF BACTERIA OF THE GENUS CHLAM
;  TITLE OF INVENTION:  USE AS A PROBE, PRIMER, AND IN A REAGENT AND A DETECTION PROCEDU
;  FILE REFERENCE: 102682
;  CURRENT APPLICATION NUMBER: US/09/230,926A
;  CURRENT FILING DATE:  1999-03-04
;  PRIOR APPLICATION NUMBER: PCT/FR98/01157
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: FR 97/07200
;  PRIOR FILING DATE: 1997-06-05
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
;   LENGTH: 47
;   TYPE: RNA
;   ORGANISM: Chlamydia pneumoniae
US-09-230-926A-35

  Query Match             40.7%;  Score 11.8;  DB 10;  Length 47;
  Best Local Similarity   65.0%;  Pred. No. 7.5e+03;
  Matches   13;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          6 UNCUUUNNGUAAGCCCNANG 25
              | ||     |||||||| | |
Db         28 UCCUCGCCGUAAGCCCAAGG 47


RESULT 11
US-09-908-975-15914/c
; Sequence 15914, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15914
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-908-975-15914

  Query Match             40.7%;  Score 11.8;  DB 13;  Length 60;
  Best Local Similarity   50.0%;  Pred. No. 7.8e+03;
  Matches    9;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCC 21
              ||  |:::  |: |||||
Db         30 GAGGCTTTGAGTGAGCCC 13


RESULT 12
US-09-908-975-17626
; Sequence 17626, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE \
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17626
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-908-975-17626

  Query Match             40.7%;  Score 11.8;  DB 13;  Length 60;
  Best Local Similarity   40.0%;  Pred. No. 7.8e+03;
  Matches    8;  Conservative    5;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          6 UNCUUUNNGUAAGCCCNANG 25
              : |:::  |:|||| |   |
Db         28 TGCTTTTGGTAAGCACTTTG 47


RESULT 13
US-09-908-975-1254
; Sequence 1254, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE \
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1254
;   LENGTH: 65
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-09-908-975-1254

  Query Match             40.7%;  Score 11.8;  DB 13;  Length 65;
  Best Local Similarity   45.0%;  Pred. No. 7.9e+03;
  Matches    9;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          6 UNCUUUNNGUAAGCCCNANG 25
              : |: :  |:|||| | | |
Db         11 TGCTCTTGGTAAGCTCCAGG 30


RESULT 14
US-09-908-975-30297
; Sequence 30297, Application US/09908975
```

```
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30297
;   LENGTH: 65
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-908-975-30297

  Query Match             40.7%;  Score 11.8;  DB 13;  Length 65;
  Best Local Similarity   50.0%;  Pred. No. 7.9e+03;
  Matches    9;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCC 21
              ||: |:::    ||||||
Db         37 GATTCTTTCCCCAAGCCC 54


RESULT 15
US-10-032-585-316
; Sequence 316, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
;  APPLICANT: Terry, Roemer D.
;  APPLICANT:  Bo, Jiang
;  APPLICANT:  Charles, Boone
;  APPLICANT:  Howard, Bussey
;  TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
;  FILE REFERENCE: 10182-005-999
;  CURRENT APPLICATION NUMBER: US/10/032,585
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 8000
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 316
;   LENGTH: 65
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-10-032-585-316

  Query Match             40.7%;  Score 11.8;  DB 13;  Length 65;
  Best Local Similarity   42.1%;  Pred. No. 7.9e+03;
  Matches    8;  Conservative    5;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          5 AUNCUUUNNGUAAGCCCNA 23
              |: |:::  |:||  || |
Db         14 ATACTTTCAGTAATACCAA 32


Search completed: January 30, 2004, 13:10:22
Job time : 182 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 06:57:47 ; Search time 575.333 Seconds
(without alignments)
2062.073 Million cell updates/sec

Title: US-09-310-844C-23
Perfect score: 29
Sequence: 1 nnngauncuuunnguaagcccnangngnn 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 1427288

Minimum DB seq length: 0
Maximum DB seq length: 70

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	12.8	44.1	53	10	MDTRVNJB	X63580 M.domesticu
	2	12.2	42.1	44	6	AX008706	AX008706 Sequence
c	3	12.2	42.1	45	6	AX008707	AX008707 Sequence
	4	12.2	42.1	69	6	AR052906	AR052906 Sequence
	5	12.2	42.1	69	6	AR054269	AR054269 Sequence
	6	12.2	42.1	69	6	AR054471	AR054471 Sequence
c	7	12.2	42.1	70	6	A42881	A42881 Sequence 13
c	8	12.2	42.1	70	6	AR271415	AR271415 Sequence
c	9	12	41.4	24	6	AX291705	AX291705 Sequence
	10	11.8	40.7	21	6	AR080204	AR080204 Sequence
	11	11.8	40.7	21	6	AX088729	AX088729 Sequence
c	12	11.8	40.7	21	6	AX088730	AX088730 Sequence
	13	11.8	40.7	21	6	BD023126	BD023126 Glutathio
c	14	11.8	40.7	33	10	MDTRVNJK	X63589 M.domesticu
c	15	11.8	40.7	40	10	MDTRVNJA	X63579 M.domesticu
c	16	11.8	40.7	46	10	MDTRVNJC	X63581 M.domesticu
	17	11.8	40.7	47	6	A82690	A82690 Sequence 35
	18	11.8	40.7	47	6	A82705	A82705 Sequence 50
c	19	11.8	40.7	51	10	MDTRVNJD	X63582 M.domesticu
	20	11.8	40.7	65	6	AX483016	AX483016 Sequence
c	21	11.8	40.7	68	7	PT7CLS	M11570 Bacteriopha
	22	11.6	40.0	24	6	AX721991	AX721991 Sequence
c	23	11.6	40.0	31	6	AX425989	AX425989 Sequence
c	24	11.6	40.0	36	6	AR176463	AR176463 Sequence
	25	11.6	40.0	36	6	AR176468	AR176468 Sequence
c	26	11.6	40.0	56	6	AX247478	AX247478 Sequence
	27	11.6	40.0	57	9	S78643	S78643 Ig VH3A10=I
	28	11.6	40.0	60	12	SYNANVAA	M60029 Avian neovi
	29	11.6	40.0	60	12	SYNANVAP	M60085 Avian neovi
c	30	11.6	40.0	61	6	AR233863	AR233863 Sequence
	31	11.6	40.0	70	6	BD107373	BD107373 DNA conta
	32	11.4	39.3	17	6	AX672857	AX672857 Sequence
	33	11.4	39.3	17	6	AX728660	AX728660 Sequence
c	34	11.4	39.3	32	17	HSMC42B09	X88068 H.sapiens D
	35	11.4	39.3	36	6	A61637	A61637 Sequence 32
c	36	11.4	39.3	40	6	AX306333	AX306333 Sequence
c	37	11.4	39.3	49	6	AX555177	AX555177 Sequence
	38	11.4	39.3	51	6	AX204359	AX204359 Sequence
c	39	11.4	39.3	70	4	BOVMTTRV	K00258 Bovine mito
	40	11.2	38.6	20	6	AX293832	AX293832 Sequence
	41	11.2	38.6	24	6	AX288322	AX288322 Sequence
	42	11.2	38.6	24	6	AX289199	AX289199 Sequence
c	43	11.2	38.6	25	6	AR020995	AR020995 Sequence
c	44	11.2	38.6	25	6	AR043410	AR043410 Sequence
c	45	11.2	38.6	25	6	AR062325	AR062325 Sequence

ALIGNMENTS

RESULT 1
MDTRVNJB/c
LOCUS MDTRVNJB 53 bp mRNA linear ROD 07-MAR-1993
DEFINITION M.domesticus DBA/2 rearranged T-cell receptor (Vgamma2-N-Jgamma2).
ACCESSION X63580
VERSION X63580.1 GI:57892
KEYWORDS joining region; N-region; T-cell receptor; variable region.
SOURCE Mus musculus domesticus (western European house mouse)
ORGANISM Mus musculus domesticus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 53)
AUTHORS Roger,T.R.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 53)

```
  AUTHORS   Roger,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1991) T. Roger, Laboratoire
            d'Immunodifferenciation, Service du Pr SEMAN, Institut J.MONOD, 2,
            Place JUSSIEU, 75251 PARIS CEDEX 05, FRANCE
FEATURES             Location/Qualifiers
     source          1. .53
                     /organism="Mus musculus domesticus"
                     /mol_type="mRNA"
                     /strain="DBA/2"
                     /db_xref="taxon:10092"
                     /chromosome="13"
                     /clone="1G3gamma5"
                     /cell_line="1G3"
                     /cell_type="CD4- CD8- T Lymphocyte clone"
                     /tissue_type="Spleen"
                     /clone_lib="library lambda gt11"
                     /dev_stage="Seed, cell expansion stage"
     misc_feature    1. .24
                     /product="End of the T-cell receptor Vgamma2 gene segment"
     misc_feature    25. .31
                     /product="N Sequence"
     misc_feature    32. .53
                     /product="Begining of the T cell receptor Jgamma2 gene
                     segment"
BASE COUNT       10 a      15 c      11 g      17 t
ORIGIN

  Query Match             44.1%;  Score 12.8;  DB 10;  Length 53;
  Best Local Similarity   52.4%;  Pred. No. 1.2e+04;
  Matches   11;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          5 AUNCUUUNNGUAAGCCCNANG 25
              |: | ::  | ||||||  | |
Db         51 ATACCTTGTGAAAGCCCGAGG 31


RESULT 2
AX008706
LOCUS       AX008706                 44 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 20 from Patent WO9965947.
ACCESSION   AX008706
VERSION     AX008706.1  GI:9996217
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Kenigsberg,M., Duchesne,M., Barlat,I. and Parker,F.
  TITLE     Monoclonal antibodies directed against the g3bp protein, and uses
  JOURNAL   Patent: WO 9965947-A 20 23-DEC-1999;
            KENIGSBERG MIREILLE (FR); DUCHESNE MARC (FR); BARLAT ISABELLE (FR);
            PARKER FABIENNE (FR); RHONE POULENC RORER SA (FR)
FEATURES             Location/Qualifiers
     source          1. .44
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="polynucleotide sq44g2s"
BASE COUNT       13 a       6 c      15 g      10 t
ORIGIN

  Query Match             42.1%;  Score 12.2;  DB 6;  Length 44;
  Best Local Similarity   54.5%;  Pred. No. 2.7e+04;
  Matches   12;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANG 25
              ||: |:    | ||||||  | |
Db         11 GATGCTAGTGGAAAGCCCCAGG 32


RESULT 3
AX008707/c
LOCUS       AX008707                 45 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 21 from Patent WO9965947.
ACCESSION   AX008707
VERSION     AX008707.1  GI:9996218
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Kenigsberg,M., Duchesne,M., Barlat,I. and Parker,F.
  TITLE     Monoclonal antibodies directed against the g3bp protein, and uses
  JOURNAL   Patent: WO 9965947-A 21 23-DEC-1999;
            KENIGSBERG MIREILLE (FR); DUCHESNE MARC (FR); BARLAT ISABELLE (FR);
            PARKER FABIENNE (FR); RHONE POULENC RORER SA (FR)
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="polynucleotide sq44g2as"
BASE COUNT        9 a      16 c       7 g      13 t
ORIGIN

  Query Match             42.1%;  Score 12.2;  DB 6;  Length 45;
  Best Local Similarity   54.5%;  Pred. No. 2.8e+04;
  Matches   12;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANG 25
              ||: |:    | ||||||  | |
Db         38 GATGCTAGTGGAAAGCCCCAGG 17


RESULT 4
AR052906
LOCUS       AR052906                 69 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 30 from patent US 5833976.
ACCESSION   AR052906
VERSION     AR052906.1  GI:5977768
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 69)
  AUTHORS   Malefyt,Rde.Waal., Howard,M., Hsu,D.-H., Ishida,H., O'Garra,A.,
            Spits,H. and Zlotnik,A.
  TITLE     Use of interleukin-10 (IL-10) to treat endotoxin- or
            superantigen-induced toxicity
  JOURNAL   Patent: US 5833976-A 30 10-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .69
                     /organism="unknown"
BASE COUNT       25 a      13 c      16 g      15 t
ORIGIN

  Query Match             42.1%;  Score 12.2;  DB 6;  Length 69;
  Best Local Similarity   43.5%;  Pred. No. 2.9e+04;
  Matches   10;  Conservative    4;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          5 AUNCUUUNNGUAAGCCCNANGNG 27
              |: | ::   :||||  | | | |
Db         11 ATGCCTTTAATAAGCTCCAAGAG 33


RESULT 5
AR054269
LOCUS       AR054269                 69 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 30 from patent US 5837232.
ACCESSION   AR054269
VERSION     AR054269.1  GI:5979846
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
```

```
REFERENCE   1  (bases 1 to 69)
  AUTHORS   De Waal Malefyt,R., Howard,M., Hsu,D.-H., Ishida,H., O'Garra,A.,
            Spits,H. and Zlotnik,A.
  TITLE     Use of an interleukin-10 antagonist to treat a B cell mediated
            autoimmune disorder
  JOURNAL   Patent: US 5837232-A 30 17-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .69
                     /organism="unknown"
BASE COUNT       25 a     13 c     16 g     15 t
ORIGIN

  Query Match             42.1%;  Score 12.2;  DB 6;  Length 69;
  Best Local Similarity   43.5%;  Pred. No. 2.9e+04;
  Matches   10;  Conservative    4;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          5 AUNCUUUNNGUAAGCCCNANGNG 27
              |: | ::   :|||| | | | |
Db         11 ATGCCTTTAATAAGCTCCAAGAG 33


RESULT 6
AR054471
LOCUS       AR054471                  69 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 30 from patent US 5837293.
ACCESSION   AR054471
VERSION     AR054471.1  GI:5980048
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 69)
  AUTHORS   De Waal Malefyt,R., Howard,M., Hsu,D.-H., Ishida,H., O'Garra,A.,
            Spits,H. and Zlotnik,A.
  TITLE     Use of interleukin-10 analogs for antagonists to treat endotoxin-
            or superantigen-induced toxicity
  JOURNAL   Patent: US 5837293-A 30 17-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .69
                     /organism="unknown"
BASE COUNT       25 a     13 c     16 g     15 t
ORIGIN

  Query Match             42.1%;  Score 12.2;  DB 6;  Length 69;
  Best Local Similarity   43.5%;  Pred. No. 2.9e+04;
  Matches   10;  Conservative    4;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          5 AUNCUUUNNGUAAGCCCNANGNG 27
              |: | ::   :|||| | | | |
Db         11 ATGCCTTTAATAAGCTCCAAGAG 33


RESULT 7
A42881/c
LOCUS       A42881                    70 bp    DNA     linear   PAT 06-MAR-1997
DEFINITION  Sequence 13 from Patent WO9502701.
ACCESSION   A42881
VERSION     A42881.1  GI:2298330
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 70)
  AUTHORS   Abken,H.J., Albert,W. and Jungfer,H.
  TITLE     METHOD OF IDENTIFYING HUMAN AND ANIMAL CELLS CAPABLE OF UNLIMITED
            PROLIFERATION OR TUMOUR FORMATION
  JOURNAL   Patent: WO 9502701-A 13 26-JAN-1995;
            BOEHRINGER MANNHEIM GMBH (DE)
COMMENT     Other publication DE 4323727 950309.
FEATURES             Location/Qualifiers
     source          1. .70
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT       23 a     11 c     19 g     17 t
ORIGIN

  Query Match             42.1%;  Score 12.2;  DB 6;  Length 70;
  Best Local Similarity   40.9%;  Pred. No. 2.9e+04;
  Matches    9;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANG 25
              ||: |:::  |:|  ||  | |
Db         70 GATCCTTTCGGTATTCCAGAAG 49


RESULT 8
AR271415/c
LOCUS       AR271415                  70 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 13 from patent US 6503706.
ACCESSION   AR271415
VERSION     AR271415.1  GI:29702833
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 70)
  AUTHORS   Abken,H.J., Albert,W. and Jungfer,H.
  TITLE     Method for identifying human and animal cells having an unlimited
            proliferation of tumor-formation potential
  JOURNAL   Patent: US 6503706-A 13 07-JAN-2003;
FEATURES             Location/Qualifiers
     source          1. .70
                     /organism="unknown"
BASE COUNT       23 a     11 c     19 g     17 t
ORIGIN

  Query Match             42.1%;  Score 12.2;  DB 6;  Length 70;
  Best Local Similarity   40.9%;  Pred. No. 2.9e+04;
  Matches    9;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANG 25
              ||: |:::  |:|  ||  | |
Db         70 GATCCTTTCGGTATTCCAGAAG 49


RESULT 9
AX291705/c
LOCUS       AX291705                  24 bp    DNA     linear   PAT 21-NOV-2001
DEFINITION  Sequence 3467 from Patent WO0179548.
ACCESSION   AX291705
VERSION     AX291705.1  GI:17053388
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Barany,F., Zirvi,M., Gerry,N.P., Favis,R. and Kliman,R.
  TITLE     Method of designing addressable array for detection of nucleic acid
            sequence differences using ligase detection reaction
  JOURNAL   Patent: WO 0179548-A 3467 25-OCT-2001;
            CORNELL RESEARCH FOUNDATION, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Hypothetical Probe Sequence"
BASE COUNT        6 a      5 c      8 g      5 t
ORIGIN

  Query Match             41.4%;  Score 12;  DB 6;  Length 24;
  Best Local Similarity   46.7%;  Pred. No. 3.5e+04;
  Matches    7;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          5 AUNCUUUNNGUAAGC 19
```

```
             |: |:::  |:||||
Db         23 ATCCTTTCCGTAAGC 9


RESULT 10
AR080204
LOCUS       AR080204                 21 bp    DNA     linear   PAT 31-AUG-2000
DEFINITION  Sequence 10 from patent US 5968737.
ACCESSION   AR080204
VERSION     AR080204.1  GI:10006939
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Ali-Osman,F., Lopez-Berestein,G., Buolamwini,J.K., Antoun,G.,
            Lo,H.-W., Keller,C. and Akande,O.
  TITLE     Method of identifying inhibitors of glutathione S-transferase (GST)
            gene expression
  JOURNAL   Patent: US 5968737-A 10 19-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unknown"
BASE COUNT        3 a      5 c      8 g      5 t
ORIGIN

  Query Match             40.7%;  Score 11.8;  DB 6;  Length 21;
  Best Local Similarity   50.0%;  Pred. No. 4.6e+04;
  Matches    9;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCC 21
              ||  |:::  |: |||||
Db          2 GAGGCTTTGAGTGAGCCC 19


RESULT 11
AX088729
LOCUS       AX088729                 21 bp    DNA     linear   PAT 17-MAR-2001
DEFINITION  Sequence 55 from Patent WO0114416.
ACCESSION   AX088729
VERSION     AX088729.1  GI:13397525
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Neeper,M.P., Mcclements,W.L., Jansen,K.U., Schultz,L.D., Chen,L.
            and Wang,X.M.
  TITLE     Synthetic human papillomavirus genes
  JOURNAL   Patent: WO 0114416-A 55 01-MAR-2001;
            Merck & Co., Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Codon-Optimized HPV16 E1 fragment"
BASE COUNT        2 a      7 c      7 g      5 t
ORIGIN

  Query Match             40.7%;  Score 11.8;  DB 6;  Length 21;
  Best Local Similarity   45.0%;  Pred. No. 4.6e+04;
  Matches    9;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          6 UNCUUUNNGUAAGCCCNANG 25
              : |: :  |: ||||| | |
Db          2 TGCTGTGCGTGAGCCCCATG 21


RESULT 12
AX088730/c
LOCUS       AX088730                 21 bp    DNA     linear   PAT 17-MAR-2001
DEFINITION  Sequence 56 from Patent WO0114416.
ACCESSION   AX088730
VERSION     AX088730.1  GI:13397526
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Neeper,M.P., Mcclements,W.L., Jansen,K.U., Schultz,L.D., Chen,L.
            and Wang,X.M.
  TITLE     Synthetic human papillomavirus genes
  JOURNAL   Patent: WO 0114416-A 56 01-MAR-2001;
            Merck & Co., Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Codon-Optimized HPV16 E1 fragment"
BASE COUNT        5 a      7 c      7 g      2 t
ORIGIN

  Query Match             40.7%;  Score 11.8;  DB 6;  Length 21;
  Best Local Similarity   45.0%;  Pred. No. 4.6e+04;
  Matches    9;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          6 UNCUUUNNGUAAGCCCNANG 25
              : |: :  |: ||||| | |
Db         20 TGCTGTGCGTGAGCCCCATG 1


RESULT 13
BD023126
LOCUS       BD023126                 21 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Glutathione S-transferase (GST) gene in cancer.
ACCESSION   BD023126
VERSION     BD023126.1  GI:22564349
KEYWORDS    JP 2001504340-A/6.
SOURCE      Wolinella succinogenes
  ORGANISM  Wolinella succinogenes
            Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
            Helicobacteraceae; Wolinella.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Aliosman,F., Berestein,G.L., Buolamwini,J.K., Antoun,G., Lo,H.W.,
            Keller,C. and Akande,O.
  TITLE     Glutathione S-transferase (GST) gene in cancer
  JOURNAL   Patent: JP 2001504340-A 6 03-APR-2001;
            BOARD OF REGENTS THE UNIVERSITY OF TEXAS SYSTEM, THE UNIVERSITY OF
            MISSISSIPPI
COMMENT     PN   JP 2001504340-A/6
            PD   03-APR-2001
            PF   12-NOV-1997 JP 1998522894
            PR   12-NOV-1996 US    08/747536
            PI   FRANCIS ALIOSMAN,GABRIEL LOPEZ BERESTEIN,JOHN K BUOLAMWINI, PI
              GAMIL ANTOUN,
            PI   HUI WEN LO,CHARLES KELLER,OLANIKE AKANDE
            PC   C12N15/09,A61K31/7105,A61K31/711,A61K38/00,A61K39/395 PC
            ,A61K39/395,A61K45/00,
            PC   A61K48/00,A61P35/00,A61P43/00,C07K16/40,C12N5/10,C12N9/00, PC
             C12N9/10,
            PC   C12Q1/02,C12N15/00,C12N5/00,A61K37/02
            CC   Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers.
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="Wolinella succinogenes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:844"
BASE COUNT        3 a      5 c      8 g      5 t
ORIGIN

  Query Match             40.7%;  Score 11.8;  DB 6;  Length 21;
  Best Local Similarity   50.0%;  Pred. No. 4.6e+04;
```

```
  Matches    9;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCC 21
              ||  |:::  |: |||||
Db          2 GAGGCTTTGAGTGAGCCC 19


RESULT 14
MDTRVNJK/c
LOCUS       MDTRVNJK                  33 bp    mRNA    linear   ROD 07-MAR-1993
DEFINITION  M.domesticus BALB/c rearranged T-cell receptor (Vgamma1-N-Jgamma4).
ACCESSION   X63589
VERSION     X63589.1  GI:57901
KEYWORDS    joining region; N-region; T-cell receptor; variable region.
SOURCE      Mus musculus domesticus (western European house mouse)
  ORGANISM  Mus musculus domesticus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 33)
  AUTHORS   Roger,T.R.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 33)
  AUTHORS   Roger,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1991) T. Roger, Laboratoire
            d'Immunodifferenciation, Service du Pr SEMAN, Institut J.MONOD, 2,
            Place JUSSIEU, 75251 PARIS CEDEX 05, FRANCE
FEATURES             Location/Qualifiers
     source          1. .33
                     /organism="Mus musculus domesticus"
                     /mol_type="mRNA"
                     /strain="BALB/c"
                     /db_xref="taxon:10092"
                     /chromosome="13"
                     /clone="TKR1902 Q"
                     /cell_line="TKR1902"
                     /cell_type="T Lymphocyte"
                     /tissue_type="Spleen"
                     /clone_lib="library M13mp18"
                     /dev_stage="Seed, cell expansion stage"
     misc_feature    1. .2
                     /product="End of the T-cell receptor Vgamma1 gene segment"
     misc_feature    3. .7
                     /product="N Sequence"
     misc_feature    8. .33
                     /product="Begining of the T cell receptor Jgamma4 gene
                     segment"
BASE COUNT        8 a      6 c     10 g      9 t
ORIGIN

  Query Match             40.7%;  Score 11.8;  DB 10;  Length 33;
  Best Local Similarity   52.6%;  Pred. No. 4.8e+04;
  Matches   10;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          5 AUNCUUUNNGUAAGCCCNA 23
              |: | ::  | ||||| |
Db         31 ATACCTTGTGAAAGCCCGA 13


RESULT 15
MDTRVNJA/c
LOCUS       MDTRVNJA                  40 bp    mRNA    linear   ROD 07-MAR-1993
DEFINITION  M.domesticus DBA/2 rearranged T-cell receptor (Vgamma2-N-Jgamma2).
ACCESSION   X63579
VERSION     X63579.1  GI:57891
KEYWORDS    joining region; N-region; T-cell receptor; variable region.
SOURCE      Mus musculus domesticus (western European house mouse)
  ORGANISM  Mus musculus domesticus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   Roger,T.R.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 40)
  AUTHORS   Roger,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1991) T. Roger, Laboratoire
            d'Immunodifferenciation, Service du Pr SEMAN, Institut J.MONOD, 2,
            Place JUSSIEU, 75251 PARIS CEDEX 05, FRANCE
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="Mus musculus domesticus"
                     /mol_type="mRNA"
                     /strain="DBA/2"
                     /db_xref="taxon:10092"
                     /chromosome="13"
                     /clone="03F9A"
                     /cell_line="03F9"
                     /cell_type="CD4+ CD8- T Lymphocyte clone"
                     /tissue_type="Spleen"
                     /clone_lib="library M13mp18"
                     /dev_stage="Seed, cell expansion stage"
     misc_feature    1. .6
                     /product="End of the T-cell receptor Vgamma2 gene segment"
     misc_feature    7. .14
                     /product="N Sequence"
     misc_feature    15. .40
                     /product="Begining of the T cell receptor Jgamma2 gene
                     segment"
BASE COUNT       10 a     11 c      8 g     11 t
ORIGIN

  Query Match             40.7%;  Score 11.8;  DB 10;  Length 40;
  Best Local Similarity   52.6%;  Pred. No. 4.9e+04;
  Matches   10;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          5 AUNCUUUNNGUAAGCCCNA 23
              |: | ::  | ||||| |
Db         38 ATACCTTGTGAAAGCCCGA 20


Search completed: January 30, 2004, 08:51:19
Job time : 582.333 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 06:19:17 ; Search time 283.333 Seconds
(without alignments)
276.295 Million cell updates/sec

Title: US-09-310-844C-23
Perfect score: 29
Sequence: 1 nnngauncuuunnguaagcccnangngnn 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 2640686

Minimum DB seq length: 0
Maximum DB seq length: 70

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	18	62.1	29	21	AAA70827	Molecular interact
2	18	62.1	29	21	AAA70828	Molecular interact
3	18	62.1	29	21	AAA70829	Molecular interact
4	18	62.1	29	21	AAA70830	Molecular interact
5	18	62.1	42	21	AAA71113	Molecular interact
6	18	62.1	42	21	AAA71114	Molecular interact
7	18	62.1	42	21	AAA71115	Molecular interact
8	18	62.1	42	21	AAA71116	Molecular interact
9	18	62.1	42	21	AAA71118	Molecular interact
10	18	62.1	42	21	AAA71119	Molecular interact
11	18	62.1	42	21	AAA71120	Molecular interact
12	18	62.1	42	21	AAA71121	Molecular interact
13	18	62.1	42	21	AAA71123	Molecular interact
14	18	62.1	42	21	AAA71124	Molecular interact
15	18	62.1	42	21	AAA71126	Molecular interact
16	18	62.1	42	21	AAA71127	Molecular interact
17	18	62.1	42	21	AAA71128	Molecular interact
18	18	62.1	42	21	AAA71129	Molecular interact
19	18	62.1	42	21	AAA71131	Molecular interact
20	18	62.1	42	21	AAA71132	Molecular interact
21	18	62.1	44	21	ABK87476	Interleukin-2 (IL-
22	18	62.1	44	21	AAA71112	Molecular interact
23	18	62.1	44	21	AAA71125	Molecular interact
24	18	62.1	44	21	AAA71133	Molecular interact
25	18	62.1	45	21	AAA70824	Molecular interact
26	18	62.1	45	21	AAA70825	Molecular interact
27	18	62.1	45	21	AAA70826	Molecular interact
28	18	62.1	46	21	AAA71085	Molecular interact
29	18	62.1	46	21	AAA71087	Molecular interact
30	18	62.1	46	21	AAA71088	Molecular interact
31	18	62.1	46	21	AAA71089	Molecular interact
32	18	62.1	46	21	AAA71090	Molecular interact
33	18	62.1	46	21	AAA71093	Molecular interact
34	18	62.1	46	21	AAA71094	Molecular interact
35	18	62.1	46	21	AAA71095	Molecular interact
36	18	62.1	46	21	AAA71096	Molecular interact
37	18	62.1	46	21	AAA71099	Molecular interact
38	18	62.1	46	21	AAA71100	Molecular interact
39	18	62.1	46	21	AAA71103	Molecular interact
40	18	62.1	46	21	AAA71104	Molecular interact
41	18	62.1	46	21	AAA71105	Molecular interact
42	18	62.1	46	21	AAA71106	Molecular interact
43	18	62.1	46	21	AAA71107	Molecular interact
44	18	62.1	46	21	AAA71109	Molecular interact
45	18	62.1	46	21	AAA71110	Molecular interact

ALIGNMENTS

RESULT 1
AAA70827
ID AAA70827 standard; RNA; 29 BP.
XX
AC AAA70827;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #27.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Synthetic.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,

PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Claim 235; Page 235; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 29 BP; 4 A; 4 C; 5 G; 5 U; 11 other;

Query Match 62.1%; Score 18; DB 21; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GAUNCUUUNNGUAAGCCCNANGNG 27
||||||||||||||||||||||||||
Db 4 GAUNCUUUNNGUAAGCCCNANGNG 27

RESULT 2
AAA70828
ID AAA70828 standard; RNA; 29 BP.
XX
AC AAA70828;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #28.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Homo sapiens.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Claim 235; Page 235; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 29 BP; 5 A; 5 C; 7 G; 12 U; 0 other;

Query Match 62.1%; Score 18; DB 21; Length 29;
Best Local Similarity 75.0%; Pred. No. 1.5;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 4 GAUNCUUUNNGUAAGCCCNANGNG 27
||| |||| |||||||| | | |
Db 4 GAUUCUUUUUGUAAGCCCUAGGGG 27

RESULT 3
AAA70829
ID AAA70829 standard; RNA; 29 BP.
XX
AC AAA70829;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #29.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Mus sp.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -

XX
PS Claim 235; Page 235; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 29 BP; 8 A; 6 C; 6 G; 9 U; 0 other;

```
  Query Match             62.1%;  Score 18;  DB 21;  Length 29;
  Best Local Similarity   75.0%;  Pred. No. 1.5;
  Matches   18;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||| ||||  ||||||||  | | |
Db          4 GAUUCUUUUUGUAAGCCCCAAGGG 27
```

RESULT 4
AAA70830
ID AAA70830 standard; RNA; 29 BP.
XX
AC AAA70830;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #30.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Rattus sp.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Claim 235; Page 235; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 29 BP; 8 A; 6 C; 6 G; 9 U; 0 other;

```
  Query Match             62.1%;  Score 18;  DB 21;  Length 29;
  Best Local Similarity   75.0%;  Pred. No. 1.5;
  Matches   18;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||| ||||  ||||||||  | | |
Db          4 GAUUCUUUUUGUAAGCCCCAAGGG 27
```

RESULT 5
AAA71113
ID AAA71113 standard; RNA; 42 BP.
XX
AC AAA71113;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #189.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.

```
XX
CC   This invention describes a novel method for identifying compounds which
CC   modulate the activity of a target biomolecule. The method uses
CC   3-dimensional representations of the biomolecule and a library of
CC   compounds and comprises (a) identifying at least one molecular
CC   interaction site of the target RNA; (b) generating in silico a virtual
CC   library of compounds predicted or calculated to interact with the
CC   molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC   representations of the target RNA with members of the virtual library of
CC   compounds to generate a hierarchy of the compounds ranked in accordance
CC   with their respective ability to form physical interactions with the
CC   molecular interaction site. The method also describes (1) RNA comprising
CC   a joined sequence of at least 24 nucleotides but not more than 70
CC   nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC   forming a first side of a first double stranded (ds) region; (b) 2
CC   nucleotides forming a first side of an internal loop region; (c) 4
CC   nucleotides forming a first side of a second ds region; (d) 4 or 5
CC   nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC   second side of the second ds region; (f) 4 nucleotides forming a second
CC   side of the internal loop region; and (g) 3 nucleotides forming a second
CC   side of the first ds region; (2) a purified and isolated RNA fragment
CC   comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC   methods and products can be used for identifying agents which modulate
CC   the activity of biomolecules, particularly RNA. Such agents can be used
CC   as pharmaceutical, agricultural or industrial compounds.
XX
SQ   Sequence 42 BP; 12 A; 7 C; 6 G; 17 U; 0 other;

  Query Match           62.1%;  Score 18;  DB 21;  Length 42;
  Best Local Similarity  75.0%;  Pred. No. 1.6;
  Matches   18;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||| ||||  |||||||| | | |
Db          7 GAUUCUUUUUGUAAGCCCUACGGG 30


RESULT 6
AAA71114
ID   AAA71114 standard; RNA; 42 BP.
XX
AC   AAA71114;
XX
DT   27-APR-2001  (first entry)
XX
DE   Molecular interaction site RNA #190.
XX
KW   Modulator; identification; molecular interaction; virtual library; ss.
XX
OS   Unidentified.
XX
PN   WO9958947-A2.
XX
PD   18-NOV-1999.
XX
PF   12-MAY-1999;   99WO-US10361.
XX
PR   12-MAY-1998;   98US-0076404.
PR   12-MAY-1998;   98US-0085092.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Ecker DJ,  Griffey R,  Crooke ST,  Sampath R,  Swayze E,  Mohan V;
PI   Hofstadler S,  McNeil J;
XX
DR   WPI; 2000-086439/07.
XX
PT   Identifying compounds which modulate activity of target biomolecules,
PT   used to provide compounds which can be used as pharmacological,
PT   agricultural and industrial compounds -
XX
PS   Example 7; Figure 122; 405pp; English.
XX
CC   This invention describes a novel method for identifying compounds which
CC   modulate the activity of a target biomolecule. The method uses
CC   3-dimensional representations of the biomolecule and a library of
CC   compounds and comprises (a) identifying at least one molecular
CC   interaction site of the target RNA; (b) generating in silico a virtual
CC   library of compounds predicted or calculated to interact with the
CC   molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC   representations of the target RNA with members of the virtual library of
CC   compounds to generate a hierarchy of the compounds ranked in accordance
CC   with their respective ability to form physical interactions with the
CC   molecular interaction site. The method also describes (1) RNA comprising
CC   a joined sequence of at least 24 nucleotides but not more than 70
CC   nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC   forming a first side of a first double stranded (ds) region; (b) 2
CC   nucleotides forming a first side of an internal loop region; (c) 4
CC   nucleotides forming a first side of a second ds region; (d) 4 or 5
CC   nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC   second side of the second ds region; (f) 4 nucleotides forming a second
CC   side of the internal loop region; and (g) 3 nucleotides forming a second
CC   side of the first ds region; (2) a purified and isolated RNA fragment
CC   comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC   methods and products can be used for identifying agents which modulate
CC   the activity of biomolecules, particularly RNA. Such agents can be used
CC   as pharmaceutical, agricultural or industrial compounds.
XX
SQ   Sequence 42 BP; 11 A; 8 C; 7 G; 16 U; 0 other;

  Query Match           62.1%;  Score 18;  DB 21;  Length 42;
  Best Local Similarity  75.0%;  Pred. No. 1.6;
  Matches   18;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||| ||||  |||||||| | | |
Db          7 GAUUCUUUUUGUAAGCCCUAGGCG 30


RESULT 7
AAA71115
ID   AAA71115 standard; RNA; 42 BP.
XX
AC   AAA71115;
XX
DT   27-APR-2001  (first entry)
XX
DE   Molecular interaction site RNA #191.
XX
KW   Modulator; identification; molecular interaction; virtual library; ss.
XX
OS   Unidentified.
XX
PN   WO9958947-A2.
XX
PD   18-NOV-1999.
XX
PF   12-MAY-1999;   99WO-US10361.
XX
PR   12-MAY-1998;   98US-0076404.
PR   12-MAY-1998;   98US-0085092.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Ecker DJ,  Griffey R,  Crooke ST,  Sampath R,  Swayze E,  Mohan V;
PI   Hofstadler S,  McNeil J;
XX
DR   WPI; 2000-086439/07.
XX
PT   Identifying compounds which modulate activity of target biomolecules,
PT   used to provide compounds which can be used as pharmacological,
PT   agricultural and industrial compounds -
XX
PS   Example 7; Figure 122; 405pp; English.
XX
CC   This invention describes a novel method for identifying compounds which
```

CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 U; 0 other;

Query Match 62.1%; Score 18; DB 21; Length 42;
Best Local Similarity 75.0%; Pred. No. 1.6;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||| ||||  |||||||| | | |
Db          7 GAUUCUUUUUGUAAGCCCCAAGGG 30
```

RESULT 8
AAA71116
ID AAA71116 standard; RNA; 42 BP.
XX
AC AAA71116;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #192.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 U; 0 other;

Query Match 62.1%; Score 18; DB 21; Length 42;
Best Local Similarity 75.0%; Pred. No. 1.6;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||| ||||  |||||||| | | |
Db          7 GAUUCUUUUUGUAAGCCCCAAGGG 30
```

RESULT 9
AAA71118
ID AAA71118 standard; DNA; 42 BP.
XX
AC AAA71118;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #124.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 125; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of

CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 12 A; 7 C; 6 G; 17 T; 0 other;

Query Match 62.1%; Score 18; DB 21; Length 42;
Best Local Similarity 54.2%; Pred. No. 1.6;
Matches 13; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

```
Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||: |::: |:|||||| | | |
Db          7 GATTCTTTTTGTAAGCCCTACGGG 30
```

RESULT 10
AAA71119
ID AAA71119 standard; DNA; 42 BP.
XX
AC AAA71119;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #125.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 125; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 11 A; 8 C; 7 G; 16 T; 0 other;

Query Match 62.1%; Score 18; DB 21; Length 42;
Best Local Similarity 54.2%; Pred. No. 1.6;
Matches 13; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

```
Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||: |::: |:|||||| | | |
Db          7 GATTCTTTTTGTAAGCCCTAGGCG 30
```

RESULT 11
AAA71120
ID AAA71120 standard; DNA; 42 BP.
XX
AC AAA71120;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #126.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 125; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual

CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 T; 0 other;

Query Match 62.1%; Score 18; DB 21; Length 42;
Best Local Similarity 54.2%; Pred. No. 1.6;
Matches 13; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

```
Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||: |::: |:|||||| | | |
Db          7 GATTCTTTTTGTAAGCCCCAAGGG 30
```

RESULT 12
AAA71121
ID AAA71121 standard; DNA; 42 BP.
XX
AC AAA71121;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #127.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 125; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 T; 0 other;

Query Match 62.1%; Score 18; DB 21; Length 42;
Best Local Similarity 54.2%; Pred. No. 1.6;
Matches 13; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

```
Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||: |::: |:|||||| | | |
Db          7 GATTCTTTTTGTAAGCCCCAAGGG 30
```

RESULT 13
AAA71123
ID AAA71123 standard; DNA; 42 BP.
XX
AC AAA71123;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #129.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 125; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)

CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 9 A; 6 C; 9 G; 18 T; 0 other;

Query Match 62.1%; Score 18; DB 21; Length 42;
Best Local Similarity 54.2%; Pred. No. 1.6;
Matches 13; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 4 GAUNCUUUNNGUAAGCCCNANGNG 27
||: |::: |:|||||| | | |
Db 7 GATTCTTTTTGTAAGCCCTAGGGG 30

RESULT 14
AAA71124
ID AAA71124 standard; DNA; 42 BP.
XX
AC AAA71124;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #130.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 125; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 11 A; 10 C; 7 G; 14 T; 0 other;

Query Match 62.1%; Score 18; DB 21; Length 42;
Best Local Similarity 54.2%; Pred. No. 1.6;
Matches 13; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 4 GAUNCUUUNNGUAAGCCCNANGNG 27
||: |::: |:|||||| | | |
Db 7 GATCCTTTCTGTAAGCCCTACGGG 30

RESULT 15
AAA71126
ID AAA71126 standard; RNA; 42 BP.
XX
AC AAA71126;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #195.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,
PT agricultural and industrial compounds -
XX
PS Example 7; Figure 126; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance

CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 12 A; 7 C; 6 G; 17 U; 0 other;

Query Match 62.1%; Score 18; DB 21; Length 42;
Best Local Similarity 75.0%; Pred. No. 1.6;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||| ||||  |||||||| | | ||
Db          7 GAUUCUUUUUGUAAGCCCUACGGG 30
```

Search completed: January 30, 2004, 08:22:12
Job time : 285.667 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 07:56:58 ; Search time 50 Seconds
(without alignments)
256.002 Million cell updates/sec

Title: US-09-310-844C-23
Perfect score: 29
Sequence: 1 nnngauncuuunnguaagcccnangngnn 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 792150

Minimum DB seq length: 0
Maximum DB seq length: 70

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	12.2	42.1	27	6	5258283-10	Patent No. 5258283
	2	12.2	42.1	69	2	US-08-410-654B-30	Sequence 30, Appl
	3	12.2	42.1	69	2	US-08-474-851-30	Sequence 30, Appl
	4	12.2	42.1	69	2	US-08-481-560-30	Sequence 30, Appl
c	5	12.2	42.1	70	4	US-08-585-593A-13	Sequence 13, Appl
	6	11.8	40.7	21	2	US-08-747-536-10	Sequence 10, Appl
c	7	11.6	40.0	36	4	US-08-218-369-7	Sequence 7, Appli
	8	11.6	40.0	36	4	US-08-218-369-15	Sequence 15, Appl
c	9	11.6	40.0	36	5	PCT-US95-03742-7	Sequence 7, Appli
	10	11.6	40.0	36	5	PCT-US95-03742-15	Sequence 15, Appl
c	11	11.6	40.0	61	4	US-09-619-213B-45	Sequence 45, Appl
c	12	11.2	38.6	25	1	US-08-741-881-28	Sequence 28, Appl
c	13	11.2	38.6	25	1	US-08-739-158-28	Sequence 28, Appl
c	14	11.2	38.6	25	2	US-08-739-167-28	Sequence 28, Appl
c	15	11.2	38.6	25	3	US-08-404-796-28	Sequence 28, Appl
c	16	11.2	38.6	25	3	US-08-931-869-28	Sequence 28, Appl
c	17	11.2	38.6	25	4	US-09-350-399-28	Sequence 28, Appl
c	18	11.2	38.6	25	4	US-09-236-140A-28	Sequence 28, Appl
	19	11.2	38.6	33	1	US-08-741-881-29	Sequence 29, Appl
	20	11.2	38.6	33	1	US-08-739-158-29	Sequence 29, Appl
	21	11.2	38.6	33	2	US-08-739-167-29	Sequence 29, Appl
	22	11.2	38.6	33	3	US-08-404-796-29	Sequence 29, Appl
	23	11.2	38.6	33	3	US-08-931-869-29	Sequence 29, Appl
	24	11.2	38.6	33	4	US-09-350-399-29	Sequence 29, Appl
	25	11.2	38.6	33	4	US-09-236-140A-29	Sequence 29, Appl
c	26	11.2	38.6	36	2	US-08-642-045B-17	Sequence 17, Appl
c	27	11.2	38.6	36	3	US-08-852-268-17	Sequence 17, Appl
	28	11.2	38.6	41	4	US-09-571-774-2	Sequence 2, Appli
	29	11.2	38.6	47	4	US-09-422-978-2842	Sequence 2842, Ap
	30	11.2	38.6	52	4	US-09-310-463-6	Sequence 6, Appli
	31	11.2	38.6	52	4	US-08-842-248A-6	Sequence 6, Appli
	32	11.2	38.6	70	3	US-09-364-380-29	Sequence 29, Appl
c	33	11	37.9	31	1	US-08-323-531-71	Sequence 71, Appl
c	34	11	37.9	31	1	US-08-198-094-71	Sequence 71, Appl
c	35	11	37.9	31	3	US-08-480-640A-119	Sequence 119, App
c	36	11	37.9	31	3	US-08-295-802-119	Sequence 119, App
c	37	11	37.9	31	3	US-08-107-794A-71	Sequence 71, Appl
c	38	11	37.9	31	3	US-08-488-237A-119	Sequence 119, App
c	39	11	37.9	31	4	US-08-375-992A-119	Sequence 119, App
c	40	11	37.9	31	4	US-08-472-679H-119	Sequence 119, App
c	41	11	37.9	31	5	PCT-US93-07424-71	Sequence 71, Appl
c	42	11	37.9	31	5	PCT-US95-02087-71	Sequence 71, Appl
	43	11	37.9	70	2	US-08-488-402A-127	Sequence 127, App
	44	11	37.9	70	2	US-08-484-552A-127	Sequence 127, App
	45	11	37.9	70	5	PCT-US96-09472-127	Sequence 127, App

ALIGNMENTS

```
RESULT 1
5258283-10
;Patent No. 5258283
;    APPLICANT: FRAZIER, MARVIN E.;MALLAVIA, LOUIS P.;SAMUEL,
;JAMES E.;BACA, OSWALD G.
;    TITLE OF INVENTION: DETECTION AND DIFFERENTIATION OF COXIELLA
;BURNETII IN BIOLOGICAL FLUIDS
;    NUMBER OF SEQUENCES: 17
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/425,856
;      FILING DATE: 23-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 927,779
;      FILING DATE: 05-NOV-1986
;      APPLICATION NUMBER: 795,207
;      FILING DATE: 05-NOV-1985
;SEQ ID NO:10:
;      LENGTH: 27
5258283-10

  Query Match           42.1%;  Score 12.2;  DB 6;  Length 27;
  Best Local Similarity  40.9%;  Pred. No. 2.6e+02;
  Matches    9;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANG 25
              | : |:::  :|||||  | |
Db          4 GGTTCTTTGAATAAGCCAAATG 25


RESULT 2
US-08-410-654B-30
; Sequence 30, Application US/08410654B
; Patent No. 5833976
; GENERAL INFORMATION:
;    APPLICANT:  Rene de Waal Malefyt
;    APPLICANT:  Di-Hwei Hsu
;    APPLICANT:  Anne O'Garra
;    APPLICANT:  Hergen Spits
;    TITLE OF INVENTION:  Use of Interleukin-10 to Treat
;    TITLE OF INVENTION:  Septic Shock
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering-Plough Corporation
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  Macintosh
;       OPERATING SYSTEM:  7.5.3
;       SOFTWARE:  Microsoft Word 5.1a
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/410,654B
;       FILING DATE:  24-MAR-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/229,854
;       FILING DATE:  19-APR-1994
;       APPLICATION NUMBER:  US 07/926,853
;       FILING DATE:  06-AUG-1992
;       APPLICATION NUMBER:  US 07/742,129
;       FILING DATE:  06-AUG-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Foulke, Cynthia L.
;       REGISTRATION NUMBER:  32,364
;       REFERENCE/DOCKET NUMBER:  DX0221KQ1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  908-298-2987
;       TELEFAX:  908-298-5388
;   INFORMATION FOR SEQ ID NO:  30:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  69 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (oligonucleotide)
US-08-410-654B-30

  Query Match             42.1%;  Score 12.2;  DB 2;  Length 69;
  Best Local Similarity   43.5%;  Pred. No. 3.2e+02;
  Matches   10;  Conservative    4;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          5 AUNCUUUNNGUAAGCCCNANGNG 27
              |: | ::   :|||| | | | |
Db         11 ATGCCTTTAATAAGCTCCAAGAG 33


RESULT 3
US-08-474-851-30
; Sequence 30, Application US/08474851
; Patent No. 5837232
;  GENERAL INFORMATION:
;     APPLICANT:  Rene de Waal Malefyt
;     APPLICANT:  Di-Hwei Hsu
;     APPLICANT:  Anne O'Garra
;     APPLICANT:  Hergen Spits
;     TITLE OF INVENTION:  Use of An Interleukin-10 Antagonist to Treat
;     TITLE OF INVENTION:  A B Cell Mediated Autoimmune Disorder
;     NUMBER OF SEQUENCES:  61
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Schering-Plough Corporation
;       STREET:  2000 Galloping Hill Road
;       CITY:  Kenilworth
;       STATE:  New Jersey
;       COUNTRY:  USA
;       ZIP:  07033
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  Macintosh
;       OPERATING SYSTEM:  7.5.3
;       SOFTWARE:  Microsoft Word 6.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/474,851
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/410,654
;       FILING DATE:  24-MAR-1995
;       APPLICATION NUMBER:  US 08/229,854
;       FILING DATE:  19-APR-1994
;       APPLICATION NUMBER:  US 07/926,853
;       FILING DATE:  06-AUG-1992
;       APPLICATION NUMBER:  US 07/742,129
;       FILING DATE:  06-AUG-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Foulke, Cynthia L.
;       REGISTRATION NUMBER:  32,364
;       REFERENCE/DOCKET NUMBER:  DX0221KQ1GD
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  908-298-2987
;       TELEFAX:  908-298-5388
;   INFORMATION FOR SEQ ID NO:  30:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  69 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (oligonucleotide)
US-08-474-851-30

  Query Match             42.1%;  Score 12.2;  DB 2;  Length 69;
  Best Local Similarity   43.5%;  Pred. No. 3.2e+02;
  Matches   10;  Conservative    4;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          5 AUNCUUUNNGUAAGCCCNANGNG 27
              |: | ::   :|||| | | | |
Db         11 ATGCCTTTAATAAGCTCCAAGAG 33


RESULT 4
US-08-481-560-30
; Sequence 30, Application US/08481560
; Patent No. 5837293
;  GENERAL INFORMATION:
;     APPLICANT:  Rene de Waal Malefyt
;     APPLICANT:  Di-Hwei Hsu
;     APPLICANT:  Anne O'Garra
;     APPLICANT:  Hergen Spits
;     TITLE OF INVENTION:  Use of Interleukin-10 to Modulate
;     TITLE OF INVENTION:  Inflammation or T-Cell Mediated
;     TITLE OF INVENTION:  Immune Function
;     NUMBER OF SEQUENCES:  61
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Schering-Plough Corporation
;       STREET:  2000 Galloping Hill Road
;       CITY:  Kenilworth
;       STATE:  New Jersey
;       COUNTRY:  USA
;       ZIP:  07033
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  Macintosh
;       OPERATING SYSTEM:  7.5.3
;       SOFTWARE:  Microsoft Word 6.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/481,560
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/410,654
;       FILING DATE:  24-MAR-1995
;       APPLICATION NUMBER:  US 08/229,854
;       FILING DATE:  19-APR-1994
;       APPLICATION NUMBER:  US 07/926,853
;       FILING DATE:  06-AUG-1992
;       APPLICATION NUMBER:  US 07/742,129
;       FILING DATE:  06-AUG-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Foulke, Cynthia L.
;       REGISTRATION NUMBER:  32,364
;       REFERENCE/DOCKET NUMBER:  DX0221KQ1GC
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  908-298-2987
;       TELEFAX:  908-298-5388
```

```
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  69 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (oligonucleotide)
US-08-481-560-30

  Query Match             42.1%;  Score 12.2;  DB 2;  Length 69;
  Best Local Similarity   43.5%;  Pred. No. 3.2e+02;
  Matches   10;  Conservative    4;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          5 AUNCUUUNNGUAAGCCCNANGNG 27
              |: | ::   :|||| | | | |
Db         11 ATGCCTTTAATAAGCTCCAAGAG 33


RESULT 5
US-08-585-593A-13/c
; Sequence 13, Application US/08585593A
; Patent No. 6503706
;  GENERAL INFORMATION:
;    APPLICANT:  ABKEN, Hinrich J
;    APPLICANT:  ALBERT, Winfried
;    APPLICANT:  JUNGFER, Herbert
;    TITLE OF INVENTION:  METHOD OF IDENTIFYING HUMAN AND ANIMAL
;    TITLE OF INVENTION:  CELLS CAPABLE OF UNLIMITED PROLIFERATION OR TUMOR
;    TITLE OF INVENTION:  FORMATION
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nikaido, Marmelstein, Murray & Oram LLP
;      STREET:  655 Fifteenth Street N.W. Suite 330
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/585,593A
;      FILING DATE:  16-JAN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP94/02307
;      FILING DATE:  13-JUL-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 43 23 727.4
;      FILING DATE:  15-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murray, Robert B.
;      REGISTRATION NUMBER:  22,980
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)638-5000
;      TELEFAX:  (202)638-4810
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  70 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-585-593A-13

  Query Match             42.1%;  Score 12.2;  DB 4;  Length 70;
  Best Local Similarity   40.9%;  Pred. No. 3.2e+02;
  Matches    9;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANG 25
              ||: |:::  |:|  ||  | |
Db         70 GATCCTTTCGGTATTCCAGAAG 49


RESULT 6
US-08-747-536-10
; Sequence 10, Application US/08747536
; Patent No. 5968737
;  GENERAL INFORMATION:
;    APPLICANT:  Ali-Osman, Francis
;    APPLICANT:  Lopez-Berestein, Gabriel
;    APPLICANT:  Buolamwini, John
;    APPLICANT:  Antoun, Gamil
;    APPLICANT:  Lo, Hui-Wen
;    APPLICANT:  Keller, Charles
;    APPLICANT:  Akande, Olanike
;    TITLE OF INVENTION:  GLUTATHIONE S-TRANSFERASE (GST) GENES IN
;    TITLE OF INVENTION:  CANCER
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/747,536
;      FILING DATE:  Concurrently Herewith
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Highlander, Steven L.
;      REGISTRATION NUMBER:  37,642
;      REFERENCE/DOCKET NUMBER:  UTXC:492
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  512/418-3000
;      TELEFAX:  512/474-7577
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-747-536-10

  Query Match             40.7%;  Score 11.8;  DB 2;  Length 21;
  Best Local Similarity   50.0%;  Pred. No. 4.2e+02;
  Matches    9;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCC 21
              ||  |:::  |: |||||
Db          2 GAGGCTTTGAGTGAGCCC 19


RESULT 7
US-08-218-369-7/c
; Sequence 7, Application US/08218369
; Patent No. 6312699
;  GENERAL INFORMATION:
;    APPLICANT:  Curiel, David T.
;    APPLICANT:  Engler, Jeffrey A.
;    TITLE OF INVENTION:  Ligands Added to Adenovirus Fiber
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  1100 Peachtree Street, Suite 2800
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
```

```
;       ZIP:  30309-4530
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/218,369
;       FILING DATE:  28-MAR-1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Pabst, Patrea L.
;       REGISTRATION NUMBER:  31,284
;       REFERENCE/DOCKET NUMBER:  IGI101
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (404) 815-6508
;       TELEFAX:  (404) 815-6555
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  36 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1..36
;       OTHER INFORMATION:  /note= "Nucleotide sequence encoding a streptavidin mimic
US-08-218-369-7

  Query Match             40.0%;  Score 11.6;  DB 4;  Length 36;
  Best Local Similarity   41.7%;  Pred. No. 6.3e+02;
  Matches   10;  Conservative    4;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||  |:::  |:  | || | | |
Db         31 GAAGCTTTAGGTGGGGCCCATGAG 8


RESULT 8
US-08-218-369-15
; Sequence 15, Application US/08218369
; Patent No. 6312699
;  GENERAL INFORMATION:
;    APPLICANT:  Curiel, David T.
;    APPLICANT:  Engler, Jeffrey A.
;    TITLE OF INVENTION:  Ligands Added to Adenovirus Fiber
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  1100 Peachtree Street, Suite 2800
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-4530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/218,369
;      FILING DATE:  28-MAR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  IGI101
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 815-6508
;      TELEFAX:  (404) 815-6555
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  36 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..36
;      OTHER INFORMATION:  /note= "Nucleotides 5 through 36 are complementary to nucl
US-08-218-369-15

  Query Match             40.0%;  Score 11.6;  DB 4;  Length 36;
  Best Local Similarity   41.7%;  Pred. No. 6.3e+02;
  Matches   10;  Conservative    4;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||  |:::  |:  | || | | |
Db         10 GAAGCTTTAGGTGGGGCCCATGAG 33


RESULT 9
PCT-US95-03742-7/c
; Sequence 7, Application PC/TUS9503742
;  GENERAL INFORMATION:
;    APPLICANT:  The UAB Research Foundation
;    TITLE OF INVENTION:  Ligands Added to Adenovirus Fiber
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/03742
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  IGI101
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 873-8794
;      TELEFAX:  (404) 873-8795
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  36 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..36
;      OTHER INFORMATION:  /note= "Nucleotide sequence
;      OTHER INFORMATION:  encoding a streptavidin mimic that binds biotin."
PCT-US95-03742-7

  Query Match             40.0%;  Score 11.6;  DB 5;  Length 36;
  Best Local Similarity   41.7%;  Pred. No. 6.3e+02;
```

```
  Matches   10;  Conservative    4;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||   |:::  |:  | || | | |
Db         31 GAAGCTTTAGGTGGGGCCCATGAG 8


RESULT 10
PCT-US95-03742-15
; Sequence 15, Application PC/TUS9503742
;  GENERAL INFORMATION:
;    APPLICANT:  The UAB Research Foundation
;    TITLE OF INVENTION:  Ligands Added to Adenovirus Fiber
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/03742
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  IGI101
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 873-8794
;      TELEFAX:  (404) 873-8795
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  36 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..36
;      OTHER INFORMATION:  /note= "Nucleotides 5 through 36
;      OTHER INFORMATION:  are complementary to nucleotides 5 through 36 of
;      OTHER INFORMATION:  Sequence ID No. 7."
PCT-US95-03742-15

  Query Match             40.0%;  Score 11.6;  DB 5;  Length 36;
  Best Local Similarity   41.7%;  Pred. No. 6.3e+02;
  Matches   10;  Conservative    4;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||   |:::  |:  | || | | |
Db         10 GAAGCTTTAGGTGGGGCCCATGAG 33


RESULT 11
US-09-619-213B-45/c
; Sequence 45, Application US/09619213B
; Patent No. 6458539
; GENERAL INFORMATION:
;  APPLICANT: Gold, Larry
;  APPLICANT:  Smith, Jonathan Drew
;  APPLICANT:  Koch, Tad
;  APPLICANT:  Golden, Mace
;  TITLE OF INVENTION: Photoselection of Nucleic Acid Ligands
;  FILE REFERENCE: NEX10-5
;  CURRENT APPLICATION NUMBER: US/09/619,213B
;  CURRENT FILING DATE:  2000-07-19
;  PRIOR APPLICATION NUMBER: 09/459,553
;  PRIOR FILING DATE: 1999-12-13
;  PRIOR APPLICATION NUMBER: 09/093,293
;  PRIOR FILING DATE: 1998-06-08
;  PRIOR APPLICATION NUMBER: 08/612,895
;  PRIOR FILING DATE: 1996-03-08
;  PRIOR APPLICATION NUMBER: 08/123,935
;  PRIOR FILING DATE: 1993-09-17
;  NUMBER OF SEQ ID NOS: 100
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 61
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Nucleic Acid
;   OTHER INFORMATION: Ligand
;   NAME/KEY: modified_base
;   LOCATION: (1)..(61)
;   OTHER INFORMATION: All T's are 5-bromouracil
US-09-619-213B-45

  Query Match             40.0%;  Score 11.6;  DB 4;  Length 61;
  Best Local Similarity   45.8%;  Pred. No. 7.1e+02;
  Matches   11;  Conservative    3;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          4 GAUNCUUUNNGUAAGCCCNANGNG 27
              ||: |: :    ||||| |   |
Db         42 GATACTATGAACAAGCCCCAGTGG 19


RESULT 12
US-08-741-881-28/c
; Sequence 28, Application US/08741881
; Patent No. 5789245
;  GENERAL INFORMATION:
;    APPLICANT:  Dubensky Jr, Thomas W
;    APPLICANT:  Polo, John M.
;    APPLICANT:  Ibanez, Carlos E.
;    APPLICANT:  Chang, Stephen M.W.
;    APPLICANT:  Jolly, Douglas J.
;    APPLICANT:  Driver, David A.
;    APPLICANT:  Belli, Barbara A.
;    TITLE OF INVENTION:  EUKARYOTIC LAYERED VECTOR INITIATION SYSTEMS
;    NUMBER OF SEQUENCES:  128
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  US
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/741,881
;      FILING DATE:  30-OCT-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  930049.423C6 / 1146.007
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  28:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  25 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-741-881-28

  Query Match             38.6%;  Score 11.2;  DB 1;  Length 25;
  Best Local Similarity   36.4%;  Pred. No. 1e+03;
  Matches    8;  Conservative    5;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          6 UNCUUUNNGUAAGCCCNANGNG 27
              : |:::  |: ||||  |   |
Db         24 TCCTTTAGGTTAGCCGTACAAG 3


RESULT 13
US-08-739-158-28/c
; Sequence 28, Application US/08739158
; Patent No. 5814482
;  GENERAL INFORMATION:
;    APPLICANT:  Dubensky Jr, Thomas W
;    APPLICANT:  Polo, John M.
;    APPLICANT:  Jolly, Douglas J.
;    APPLICANT:  Driver, David A.
;    TITLE OF INVENTION:  EUKARYOTIC LAYERED VECTOR INITIATION SYSTEMS
;    NUMBER OF SEQUENCES:  128
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  US
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/739,158
;      FILING DATE:  30-OCT-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  930049.423D3 / 1146.012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  25 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-739-158-28

  Query Match             38.6%;  Score 11.2;  DB 1;  Length 25;
  Best Local Similarity   36.4%;  Pred. No. 1e+03;
  Matches    8;  Conservative    5;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          6 UNCUUUNNGUAAGCCCNANGNG 27
              : |:::  |: ||||  |   |
Db         24 TCCTTTAGGTTAGCCGTACAAG 3


RESULT 14
US-08-739-167-28/c
; Sequence 28, Application US/08739167
; Patent No. 5843723
;  GENERAL INFORMATION:
;    APPLICANT:  Dubensky Jr, Thomas W
;    APPLICANT:  Polo, John M.
;    APPLICANT:  Ibanez, Carlos E.
;    APPLICANT:  Chang, Stephen M.W.
;    APPLICANT:  Jolly, Douglas J.
;    APPLICANT:  Driver, David A.
;    APPLICANT:  Belli, Barbara A.
;    TITLE OF INVENTION:  EUKARYOTIC LAYERED VECTOR INITIATION SYSTEMS  AND ALPHAVIRUS
;    NUMBER OF SEQUENCES:  128
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  US
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/739,167
;      FILING DATE:  30-OCT-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  930049.423C7 / 1146.008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  25 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-739-167-28

  Query Match             38.6%;  Score 11.2;  DB 2;  Length 25;
  Best Local Similarity   36.4%;  Pred. No. 1e+03;
  Matches    8;  Conservative    5;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          6 UNCUUUNNGUAAGCCCNANGNG 27
              : |:::  |: ||||  |   |
Db         24 TCCTTTAGGTTAGCCGTACAAG 3


RESULT 15
US-08-404-796-28/c
; Sequence 28, Application US/08404796
; Patent No. 6015686
;  GENERAL INFORMATION:
;    APPLICANT:  Dubensky Jr, Thomas W
;    APPLICANT:  Polo, John M.
;    APPLICANT:  Ibanez, Carlos E.
;    APPLICANT:  Chang, Stephen M.W.
;    APPLICANT:  Jolly, Douglas J.
;    APPLICANT:  Driver, David A.
;    APPLICANT:  Belli, Barbara A.
;    TITLE OF INVENTION:  EUKARYOTIC LAYERED VECTOR INITIATION SYSTEMS
;    NUMBER OF SEQUENCES:  128
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  US
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/404,796
;       FILING DATE:  15-MAR-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  McMasters, David D.
;       REGISTRATION NUMBER:  33,963
;       REFERENCE/DOCKET NUMBER:  930049.423C5 / 1146.006
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  28:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  25 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-404-796-28

  Query Match             38.6%;  Score 11.2;  DB 3;  Length 25;
  Best Local Similarity   36.4%;  Pred. No. 1e+03;
  Matches    8;  Conservative    5;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          6 UNCUUUNNGUAAGCCCNANGNG 27
              : |:::  |: ||||  |   |
Db         24 TCCTTTAGGTTAGCCGTACAAG 3


Search completed: January 30, 2004, 10:15:10
Job time : 55 secs
```